DISSERTATION | DOCTORAL THESIS

Titel | Title

Prognostic and diagnostic biomarkers in clear cell renal cell carcinoma

verfasst von | submitted by

Gustav Jonsson

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 620 490
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von | Supervisor:	Dr. Josef-Martin Penninger

List of publications

Publications from the Penninger lab included in this thesis:

1. Gustav Jonsson, Maura Hofmann, Stefan Mereiter, Lauren Hartley-Tassell, Irma Sakic, Tiago Oliveira, David Hoffmann, Maria Novatchkova, Alexander Schleiffer, Josef M Penninger. "CLEC18A interacts with sulfated GAGs and controls clear cell renal cell carcinoma progression". **bioRxiv** (2024). https://doi.org/10.1101/2024.07.08.602586.

2. Gustav Jonsson, Maura Hofmann, Tiago Oliveira, Ursula Lemberger, Karel Stejskal, Gabriela Krssakova, Irma Sakic, Maria Novatchkova, Stefan Mereiter, Gerlinde Grabmann, Thomas Koecher, Zeljko Kikic, Gerald N Rechberger, Thomas Zuellig, Bernhard Englinger, Manuela Schmidinger, Josef M Penninger. "Urinary multi-omics reveal non-invasive diagnostic biomarkers in clear cell renal cell carcinoma". **bioRxiv** (2024). https://doi.org/10.1101/2024.08.12.607453.

Additional publications from the Penninger lab which are not included in this thesis:

1. Philipp Starkl, Gustav Jonsson, Tyler J Artner, Bruna Lenfers Turnes, Nadine Serhan, Tiago Oliveira, Laura-Marie Gail, Karel Stejskal, Keith Channon, Thomas Kocher, Georg Stary, Victoria Klang, Nicolas Gaudenzio, Sylvia Knapp, Clifford J Woolf, Josef M Penninger, Shane Joseph Francis Cronin. "Mast cell-derived BH4 is a critical mediator of postoperative pain." **Science Immunology** (2024). https://doi.org/10.1126/sciimmunol.adh0545.

2. Stephanie Frenz-Wiessner, Savannah D Fairley, Maximilian Buser, Isabel Goek, Kirill Salewskij, Gustav Jonsson, David Illig, Benedicta Zu Putlitz, Daniel Petersheim, Yue Li, Pin-Hsuan Chen, Martina Kalauz, Raffaele Conca, Michael Sterr, Johanna Geuder, Yoko Mizoguchi, Remco TA Megens, Monika I Linder, Daniel Kotlarz, Martina Rudelius, Josef M Penninger, Carsten Marr, Christoph Klein. "Generation of complex bone marrow organoids from human induced pluripotent stem cells." **Nature Methods** (2024). https://doi.org/10.1038/s41592-024-02172-2.

3. Clément Quintard, Emily Tubbs, Gustav Jonsson, Jie Jiao, Jun Wang, Nicolas Werschler, Camille Laporte, Amandine Pitaval, Thierno-Sidy Bah, Gideon Pomeranz, Caroline Bissardon, Joris Kaal, Alexandra Leopoldi, David A. Long, Pierre Blandin, Jean-Luc Achard, Christophe Battail, Astrid Hagelkruys, Fabrice Navarro, Yves Fouillet, Josef M. Penninger, Xavier Gidrol.

"A microfluidic platform integrating functional vascularized organoids-on-chip." **Nature Communications** (2024). https://doi.org/10.1038/s41467-024-45710-4.

4. Vanessa Monteil, Hyesoo Kwon, Lijo John, Cristiano Salata, Gustav Jonsson, Sabine U. Vorrink, Sofia Appelberg, Sonia Youanna, Matheus Dyczynski, Alexandra Leopoldi, Nicole Leeb, Jennifer Volz, Astrid Hagelkruys, Max Kellner, Stéphanie Devignot, Friedemann Weber, Volker Lauschke, Moritz Horn, Heinz Feldmann, Ulrich Elling, Josef Penninger, Georg Michlits, Ali Mirazami. "Identification of CCZ1 as an essential lysosomal trafficking regulator in filovirus infection." **Nature Communications** (2023). https://doi.org/10.1038/s41467-023-42526-6.

5. Stefan Mereiter, Gustav Jonsson, Tiago Oliveira, Johannes Helm, David Hoffmann, Markus Abeln, Ann-Kristin Jochum, Wolfram Jochum, Max J Kellner, Marek Feith, Vanessa Tkalec, Karolina Wasilewska, Jie Jiao, Lukas Emsenhuber, Felix Holstein, Anna C Obenauf, Leonardo Lordello, Jean-Yves Scoazec, Guido Kroemer, Laurence Zitvogel, Omar Hasan Ali, Lukas Flatz, Rita Gerardy-Schahn, Anja Münster-Kühnel, Johannes Stadlmann, Josef M Penninger. "Tumor sialylation controls effective anti-cancer immunity in breast cancer." **bioRxiv** (2023). https://doi.org/10.1101/2023.09.20.558571.

6. Elena Garreta, Daniel Moya-Rull, Megan L Stanifer, Vanessa Monteil, Patricia Prado, Andrés Marco, Carolina Tarantino, Maria Gallo, Gustav Jonsson, Astrid Hagelkruys, Ali Mirazimi, Steeve Boulant, Josef M Penninger, Nuria Montserrat. "Protocol for SARS-CoV-2 infection of kidney organoids derived from human pluripotent stem cells." **STAR protocols** (2022). https://doi.org/10.1016/j.xpro.2022.101872.

7. Elena Garreta, Praticia Prado, Megan L. Stanifer, Vanessa Monteil, Carmen Hurtado del Pozo, Asier Ullate-Agote, Amaia Vilas-Zornoza, Juan Pablo Romero, Gustav Jonsson, Roger Oria, Alexandra Leopoldi, Astrid Hagelkruys, Daniel Moya-Rull, Federico Gonzalez, Andres Marco, Carolina Tarantino, Pere Domingo-Pedrol, Omar Hasan Ali, Pedro Ventura-Aguiar, Josep Maria Campistol, Felipe Prosper, Ali Mirazimi, Steeve Boulant, Josef Penninger, Nuria Montserrat. "A diabetic milieu increases cellular susceptibility to SARS-CoV-2 infections in engineered human kidney organoids and diabetic patients." **Cell Metabolism** (2022). https://doi.org/10.1016/j.cmet.2022.04.009.

8. Carlos Gomez-Diaz, Gustav Jonsson, Katrin Schodl, Luiza Deszcz, Annika Bestehorn, Kevin Eislmayr, Jorge Almagro, Anoop Kavirayani, Lilian M Fennell, Astrid Hagelkruys, Pavel Kovarik, Josef M Penninger, Fumiyo Ikeda. "The ubiquitin ligase HOIL-1L regulates immune responses by interacting with linear ubiquitin chains." **iScience** (2021). https://doi.org/10.1016/j.isci.2021.103241.

9. David Hoffmann*, Stefan Mereiter*, Yoo Jin Oh, Dr. Vanessa Monteil, Elizabeth Elder, Rong Zhu, Daniel Canena, Lisa Hain, Elisabeth Laurent, Clemens Gruenwald-Gruber, Dr. Miriam Klausberger, Gustav Jonsson, Max Kellner, Maria Novatchkova, Melita Ticevic, Antoine Chabloz, Gerald Wirnsberger, Astrid Hagelkruys, Friedrich Altmann, Lukas Mach, Johannes Stadlmann, Chris Oostenbrink, Ali Mirazimi, Peter Hinterdorfer, Josef Penninger. "Identification of lectin receptors for conserved SARS-CoV-2 glycosylation sites." **The EMBO Journal** (2021). https://doi.org/10.15252/embj.2021108375.

10. Sophie Curio*, Gustav Jonsson*, Sonja Marinovic* and the young European Federation of Immunological Societies (yEFIS). "A summary of current NKG2D-based CAR clinical trials". **Immunotherapy Advances** (2021). https://doi.org/10.1093/immadv/ltab018. * = Equal contribution.

11. Astrid Hagelkruys, Gerald Wirnsberger, Johannes Stadlmann, Miriam Wöhner, Bojan Vilagos, Marion Horrer, Gustav Jonsson, Melanie Kogler, Luigi Tortola, Maria Novatchkova, Peter Bönelt, David Hoffmann, Rubina Koglgruber, Meinrad Busslinger, Andreas Bergthaler, Christoph Klein, Josef M. Penninger. "A crucial role for Jagunal homolog 1 in humoral immunity and antibody glycosylation in mice and humans". **Journal of Experimental Medicine** (2020). https://doi.org/10.1084/jem.20200559.

Acknowledgements

As I am approaching the end of my PhD, it has become evident that I could not have completed this journey without the overwhelming support that I have received from lab members, collaborators, friends and family. First and foremost, I would like to extend my deepest gratitude to my supervisor and mentor **Josef Penninger** for giving me the opportunity and trust to independently pursue the research ideas I came up with, and to establish my own line of research in the lab. Without your guidance, especially as projects were in difficult phases, this thesis would not have come together.

A massive thank you also goes out to all members of the Penninger lab, both present and past, that I have overlapped with. Special thanks go to: Team glyco, **Stefan Mereiter** and **Tiago Oliveira**, for supporting and discussing every project I have been involved in over countless coffee breaks (they are important!). And for attempting to teach me some things about glycobiology. As much as I appreciate it, I hope I never have to touch glycobiology ever again – it is truly terrible. **Astrid Hagelkrüys**, for immense support and guidance from day 1, and for always trying to solve every problem I have ever been confronted with. My master students, **Maura Hofmann** and **Irma Sakic,** without them neither one of the publications discussed in this thesis would have come together. **Shane Cronin**, for pushing me onto immunology projects and making fun of the way I pronounce G and J. I promise I will never learn. **Kirill Salewskij**, for making me laugh and the support during that one horrible kidney experiment which never will see the light of day. You know which one. **Alexandra Leopoldi**, for teaching me everything about organoids and for spending countless months cutting them out of Matrigel with me. **David Hoffmann**, for the lectin library and helping me get the CLEC18A project off the ground.

During the PhD, I have also had incredible collaborators: **Manuela Schmidinger**, **Bernhard Englinger** and **Ursula Lemberger** at AKH who assembled the entire clinical cohort. **Thomas Züllig** and **Gerhard Rechberger** in Graz who did all of the lipidomics. **Lauren Hartley-Tassell** in Australia for performing the glycoarray. And of course, all of our incredible in-house facilities without which we would not have been able to perform a fraction of the experiments we have done; Ethical approvals to recruit the urine cohort, Proteomics, Metabolomics and Bioinformatics. Special thanks go out to: **Arabella Meixner**, **Karel Stejskal**, **Gabriela Krššáková**, **Gerlinde Grabmann**, **Thomas Köcher**, **Maria Novatchkova** and **Alexander Schleiffer**. I would also like to thank my thesis advisory committee, **Anna Obenauf** and **Wilfried Ellmeier**, for their continuous support throughout my PhD journey.

I am very lucky to have formed some incredible friendships during my time in Vienna. A special shoutout goes to **Annika**, **Bianca**, **Carlos**, **Farja**, **Issy**, **Kai**, **Kathi**, **Laurenz**, **Mosti**, **Oriana**, **Sabrina**, **Saku**, **Sonya** and **Viki**. I love you all dearly and thank you so much for all of the spectacular memories – most of which are incredibly inappropriate and will never be spoken of.

I would not be where I am without my family. Thank you, **Mom** and **Dad**, for your continuous support as I decided to pack my bags and started moving across the world almost 10 years ago. You have taught me so many valuable lessons regarding strength, perseverance and standing up for yourself. Thanks to my brother, **Oscar**, for being such an important and fun part of my life. Please move a bit closer to Vienna, I am sure these parts of Europe could benefit from a some more fire safety!

Lastly, but certainly not least, a massive thank you goes out to my partner, **Markus**. Without your continuous support, especially in times when I did not believe in myself, this thesis would not have been completed. Apologies for exposing you to more science complaining than any person realistically should ever have to put up with. I love you more than I can express in a small paragraph in a downloadable PDF on a university repository that no one except for my examiners will ever read, and I look forward to spending many more years with you. And of course, big hugs to **Donna**, the best puppy in the world.

Table of Contents

Abstract

Cancer is one of the leading causes of death worldwide and constitutes a sixth of all deaths globally. However, due to breakthroughs in cancer therapy, such as immunotherapy, the overall survival rates are increasing. To further promote cancer survival, better early-stage diagnostic markers are necessary to detect the disease at an early stage to facilitate successful treatments. Furthermore, better prognostic markers are needed to predict the disease outcomes and to better stratify patients into different intervention groups. This thesis aims to discover prognostic and diagnostic biomarkers for clear cell renal cell carcinoma (ccRCC), the most common type of kidney cancer.

For prognosis, a previously poorly characterized gene family was studied: the CLEC18 gene family of C-type lectins. The CLEC18 gene family was found to be highly conserved across the entire Chordata phylum and its expression is predominantly localized to the proximal tubule of the kidney. Furthermore, we identified the ligand - sulfated glycosaminoglycans anchored to proteoglycans, a unique ligand for C-type lectins. Functionally, CLEC18 was found to inhibit ccRCC growth. In human patients, CLEC18 expression levels positively correlated with survival of ccRCC, making it a putative prognostic marker of disease outcome. These results were also validated in mouse models. For diagnosis, we performed urine proteomics, metabolomics and lipidomics from kidney cancer patient and identified three proteins with diagnostic power in ccRCC patients: Serum amyloid A1 (SAA1), Haptoglobin (HP) and Lipocalin 15 (LCN15). Remarkably, when combining SAA1, HP and LCN15 into a cumulative UrineScore, an overall performance accuracy of 96% was achieved in distinguishing between healthy controls and ccRCC patients. The UrineScore further showed potential in differentiating between ccRCC and other types of renal malignancy.

In summary, my PhD work has identified potential prognostic and diagnostic biomarkers for ccRCC.

Zusammenfassung

Krebs ist weltweit eine der häufigsten Todesursachen und macht weltweit ein Sechstel aller Todesfälle aus. Erfreulicherweise nehmen die Gesamtüberlebensraten zu. Um das Überleben von Krebs weiter zu verbessern, werden bessere diagnostische Marker für die Frühdiagnose benötigt, um die Krankheit in einem frühen Stadium zu erkennen und eine erfolgreiche Behandlung zu ermöglichen. Darüber hinaus werden bessere prognostische Marker gebraucht, um den Krankheitsverlauf vorherzusagen und die Patienten besser in verschiedene Behandlungsgruppen einteilen zu können. Ziel dieser Arbeit war die Entdeckung prognostischer und diagnostischer Biomarker für das klarzellige Nierenzellkarzinom (ccRCC), die häufigste und tödlichste Form von Nierenkrebs.

Für die Prognose wurde eine bisher wenig charakterisierte Genfamilie untersucht: die CLEC18-Genfamilie der C-Typ-Lektine. Es wurde festgestellt, dass die CLEC18-Genfamilie im gesamten Stamm der Chordata hoch konserviert ist und ihre Expression vorwiegend im proximalen Tubulus der Niere lokalisiert ist. Darüber hinaus war der Ligand von CLEC18 bisher nicht charakterisiert, aber es wurde festgestellt, dass es sich um sulfatierte Glykosaminoglykane handelt, die an Proteoglykanen verankert sind, ein einzigartiger Ligand für C-Typ-Lektine. In Bezug auf Krebs wurde festgestellt, dass CLEC18 spezifisch das Wachstum von ccRCCs hemmt und dass sein Expressionsniveau positiv mit dem Überleben korreliert, was es zu einem möglichen prognostischen Marker für den Krankheitsverlauf macht.

Für die Diagnose identifizierte die Proteomik im Urin von ccRCC-Patienten drei Proteine mit diagnostischer Aussagekraft: Serum-Amyloid A1 (SAA1), Haptoglobin (HP) und Lipocalin 15 (LCN15). Bei der Kombination von SAA1, HP und LCN15 zu einem kumulativen UrineScore wurde eine Gesamtgenauigkeit von 96% bei der Unterscheidung zwischen gesunden Kontrollen und ccRCC-Patienten erreicht. Der UrineScore zeigte darüber hinaus Potenzial bei der Unterscheidung zwischen ccRCC und anderen Arten von Nierenmalignität.

Zusammenfassend lässt sich sagen, dass meine Doktorarbeit potenzielle prognostische und diagnostische Biomarker für ccRCC identifiziert hat.

1. Introduction

The kidneys are the main filtration hubs of the human body, and they are responsible for reabsorption and secretion of solutes via the nephron [1]. In this section I will introduce the overall structure and function of the kidney, describe the function of the nephron in more detail and how regions of the nephron act as the origin for renal cell carcinomas. Furthermore, the different types of kidney cancer will be discussed with a main focus on clear cell renal cell carcinoma (ccRCC), and the current standard practices for treatment and diagnosis.

1.1 Structure and function of the kidney

Secretion of waste products is an essential function of the human body to maintain a healthy steady state homeostasis. One of the body's main secretory systems is the urinary system which consists of two kidneys, ureters, the bladder, and the urethra [2]. The functional tissue of the kidney, the parenchyma, is divided into two parts, the medulla and the cortex, and surrounded by the renal capsule (**Fig. 1**). The cortex and medulla of the kidney contains the nephrons which are the main functional units and hence responsible for secretion and resorption of solutes between blood and urine [3].

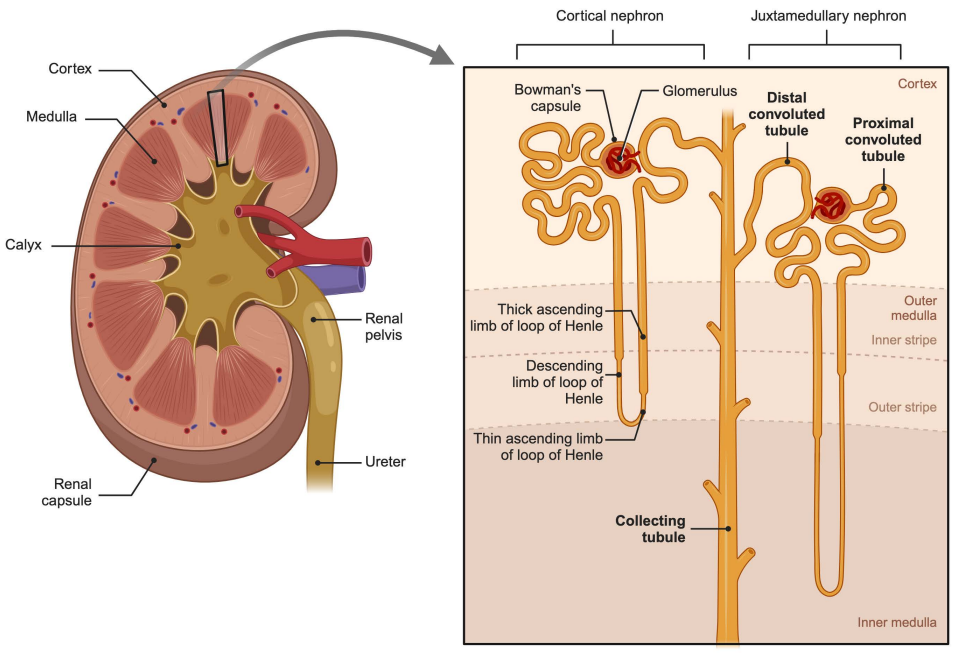


Figure 1. Anatomical structure of the human kidney. Cartoon of the gross anatomical structure of the kidney (left). Magnified cross section of a nephron and its associated structures (right). Figure created with BioRender.com.

Concentrated urine leaves the nephron and enters the collecting tubule which further transports the urine into the calyx via a porous structure called the cribriform plate [4]. Multiple calices converge into the renal pelvis and through the renal pelvis the urine effectively leaves the kidneys and is transported via the ureter into the bladder.

1.2 The nephron

On average, each adult human kidney contains 1 million nephrons. Each nephron consists of the glomerulus surrounded by the Bowman's capsule and the subsequent proximal tubules (**Fig. 1-2**) [5]. The nephron excretes waste products in the form of urine, which is formed in a multi-step process consisting of (i) filtration of the blood, (ii) reabsorption and (iii) secretion of additional solutes across the tubules and loop of Henle (**Fig. 2**) [6, 7].

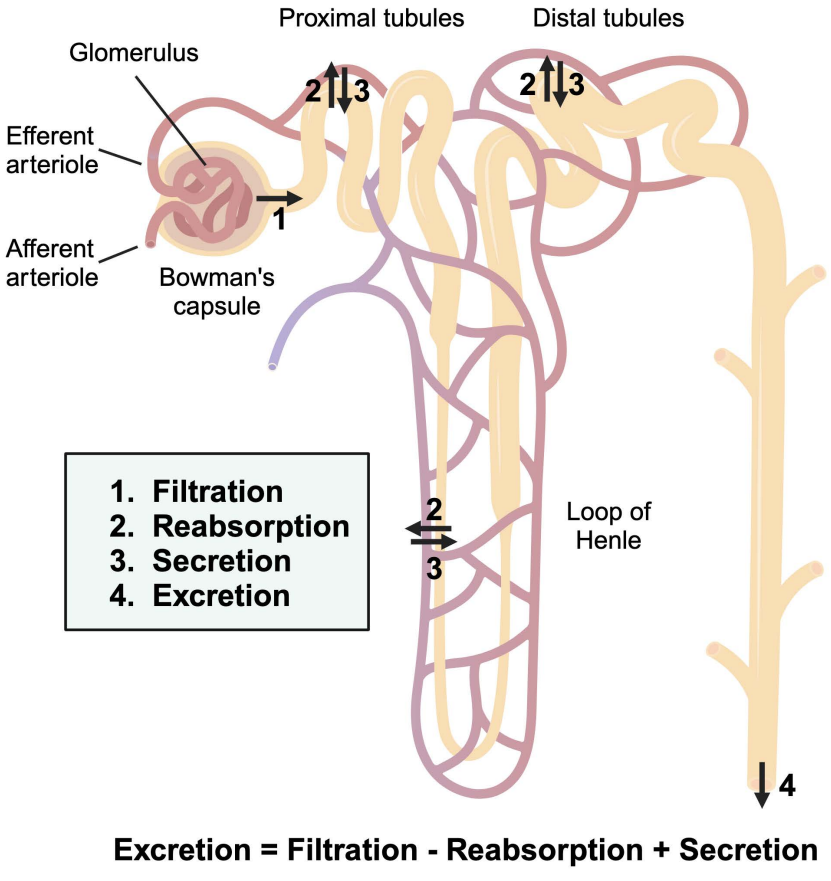


Figure 2. The excretion process of urine across the nephron. Schematic of the process through which urine is excreted from the nephron. Excretion consists of filtration of the blood, and subsequent reabsorption and secretion over the tubules and Loop of Henle. Figure created with BioRender.com.

The glomerulus is a network of small capillaries which performs the first step of the excretion process; the filtration of the blood which enters the glomerulus through the afferent arteriole. The glomerular filtration barrier selectively filters urea, salts, water, amino acids and other solutes while at the same time retaining larger proteins [8, 9]. Filtered solutes are collected in the Bowman's capsule and enter the proximal tubule [5]. Maintenance of functioning glomerular filtration and a homeostatic glomerular filtration rate (GFR) is essential for efficient clearance of waste products. As the blood enters the glomeruli, the GFR is the fluid flow rate between the glomerulus and the Bowman's capsule, which is determined by the hydrostatic and osmotic pressures in the glomerulus and the Bowman's capsule:

$$\mathrm{GFR} = \frac{dV}{dt} = K_f[P_G - P_B - (\Pi_\mathrm{G} + \Pi_\mathrm{B})]$$

where *dV/dt* is the derivative of volume with respect to time, K_f is a filtration constant, P_G and P_B are the hydrostatic pressures within the glomerular capillaries and Bowman's capsule, respectively, and $\prod_\mathrm{G}$ and $\prod_\mathrm{B}$ are the osmotic pressures within the glomerular capillaries and Bowman's capsule, respectively [10, 11].

Following filtration, the filtrate enters the proximal tubule where a majority of all reabsorption takes place. The proximal tubule actively reabsorbs salts and nutrients, such as amino acids, and passively recovers e.g. bicarbonate and water. Morphologically, the proximal tubule can be further subdivided into the proximal convoluted tubule and the proximal straight tubule, with most of the reabsorption and secretion taking place in the proximal convoluted tubule [12-14]. The luminal side of the proximal tubule epithelium has a pronounced brush border with numerous microvilli giving the luminal proximal tubule a large surface area effectively facilitating the high efficiency of reabsorption taking place in this part of the nephron [15].

Whereas the proximal tubules are predominately located in the cortex of the kidney, the subsequent part after the nephron, the loop of Henle, can extend deep into the medulla. The loop of Henle can be subdivided into the descending limb which is stretching into the medulla, and the subsequent thin and thick ascending limbs. The thin segments of the loop of Henle are permeable to water which leads to the establishment of a medullary concentration gradient throughout the loop of Henle which promotes passive secretion of salts and other solutes [16-19].

When the tubular liquid leaves the loop of Henle, it enters the distal convoluted tubule. The initial part of the distal convoluted tubule is impermeable to water leading to the formation of

a highly diluted tubular liquid as solutes are continuously reabsorbed from the lumen [20, 21]. As the tubular liquid progresses through the distal convoluted tubule and into the collecting ducts, a high volume of water is reabsorbed through aquaporins leading to the formation of concentrated urine fluid [22, 23].

A brief summary of the filtration, reabsorption and secretion steps of the urine excretion process in the glomeruli is depicted below in **Figure 3**.

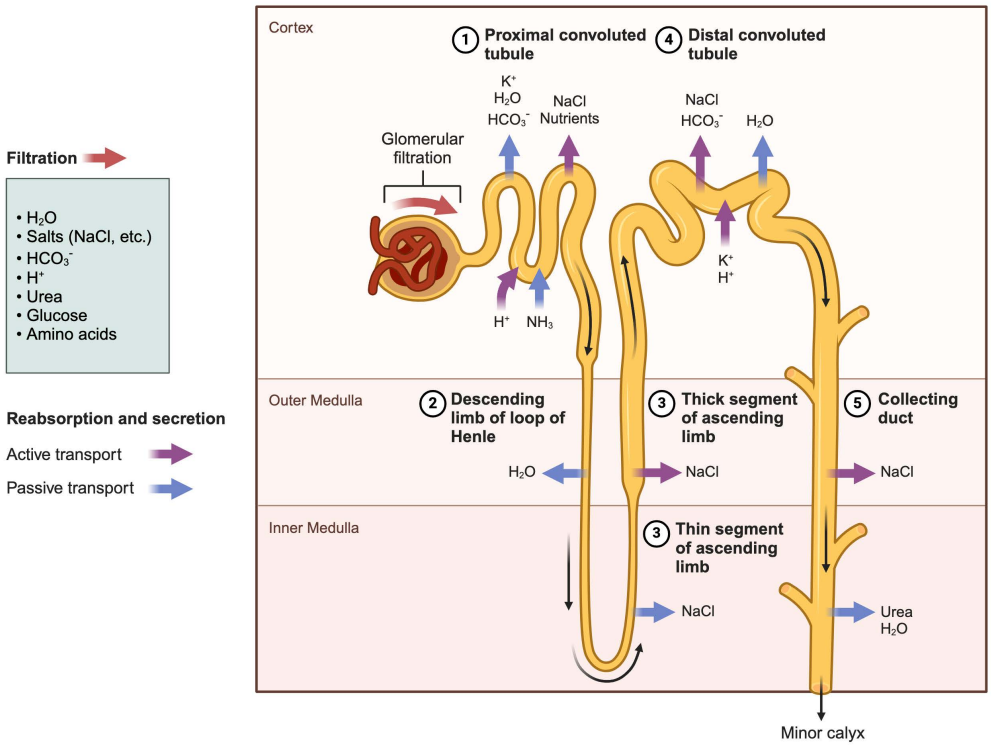


Figure 3. Filtration, reabsorption and secretion in the nephron. Schematic of glomerular filtration and reabsorption and secretion in the proximal tubule, loop of Henle, and distal convoluted tubule in the nephrons of the kidney. Created with BioRender.com.

As described, the nephron is the functional unit of the kidney and serves as the site of filtration, reabsorption and secretion, as urine is formed and expunged into the bladder. Thus, an intact and functioning nephron is essential for homeostasis. The most common form of kidney cancer, renal cell carcinoma (RCC), arises from various part of the nephron, depending on the subtype of RCC.

1.3 Renal cell carcinoma (RCC)

RCCs is the most common type of kidney cancer and constitutes more than 90% of all kidney cancer cases. The cell of origin for RCC is believed to be the renal epithelium in the nephron

[24] and RCC is predominately driven by loss of function of the *Von Hippel Lindau (VHL)* tumor suppressor [25, 26]. RCC is a very heterogenous group of cancers and more than 10 different distinct histological and molecular subtypes have been reported and associate with various clinical outcomes and disease severity [24, 27].

The most common subtype of RCC is clear cell renal cell carcinoma (ccRCC) constituting approximately 75-80% of all RCC cases [24, 28]. Histologically, ccRCC typically presents with large clear cells due to a cytoplasm with abnormally high lipid and glycogen levels [29] (**Fig. 4**). ccRCCs are predominantly derived from the proximal convoluted tubule. Multiple other non-clear cell renal cell carcinomas have been described, the two most common being papillary renal cell carcinoma (pRCC) and chromophobe renal cell carcinoma (chRCC). pRCC constitutes approximately 10-15% of all RCC cases and arises from the distal convoluted tubule. pRCC is further divided into two distinct histological subtypes; (i) Type 1 which presents with small basophilic cells of a low nuclear grade, and (ii) Type 2 which presents with eosinic cells of a high nuclear grade [30, 31] (**Fig. 4**). Lastly, chRCC makes up 3-5% of all RCC cases and the origin is the cortical collecting duct. Histologically, chRCC presents with mixed pale to eosinophilic cytoplasm [32, 33] (**Fig. 4**). Many other, less common, subtypes of RCC exist – such as Fumarate hydratase (FH) deficient RCC [34], translocation RCC [35], collecting duct RCC [36], medullary RCC [37], and mucinous tubular and spindle cell RCC [38]. These rare subtypes will not be further discussed in this thesis.

The major histological subtypes, ccRCC, pRCC and chRCC, present with unique gene profiles and phenotypes. For example, chRCCs commonly present with loss of 7 chromosomes (namely chromosomes 1, 2, 6, 10, 13, 17 and 21) [39] which is not found in other types of RCC. pRCCs on the other hand, commonly have mutations in MET and PBRM1 which are associated with poor survival (Type 1) or CDKN2A silencing, SETD2 mutations and TFE3 fusions (Type 2). Type 2 pRCCs exhibit an enhanced metabolic signature relating to glycolysis, ribose metabolism and Krebs cycle compared to Type 1 pRCCs [31, 40]. ccRCCs, just like pRCCs, commonly display PBRM1 and SET2D alterations as co-driving tumor mutations [41] in addition to loss of VHL. Furthermore, compared to other RCCs, ccRCCs have a high degree of immune infiltration [40, 42]. Despite the different genotypes and phenotypes of RCCs, DNA hypermethylation and elevated T_H2 signatures are associated with a worse prognosis for all subtypes of RCCs [40].

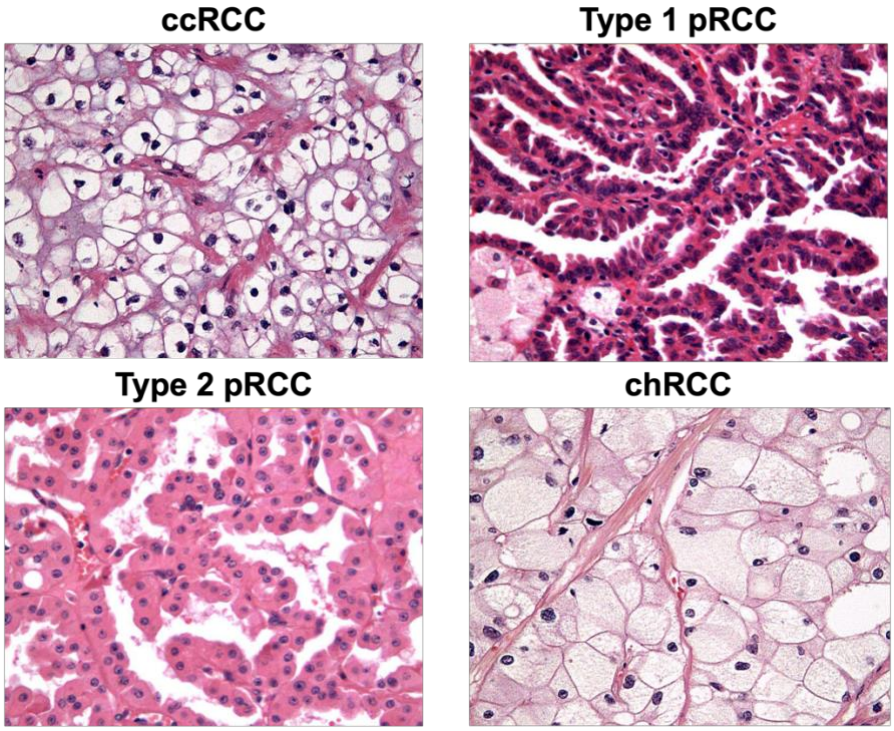


Figure 4. Representative histology from different subtypes of RCC. Examples of typical H&E staining from the four main subtypes of RCC: (i) ccRCC (top left), (ii) Type 1 pRCC (top right), (iii) Type 2 pRCC (bottom left) and (iv) chRCC (bottom right). Adapted from Algaba, F et al, European Urology, 2011 [27].

Out of these different subtypes, as previously mentioned, ccRCC is the most common type of RCC and also poses the largest burden on the healthcare system.

1.3.1 Clear cell renal cell carcinoma

Out of the three main subtypes of RCCs, ccRCC is not only the most common, but also the deadliest renal tumor [24, 28, 43]. ccRCC is primarily driven by loss-of-function mutations or epigenetic silencing of the *VHL* tumor suppressor, leading to the initiation of a genetic hypoxia program which drives tumor progression [44, 45]. Under normal oxygen conditions, VHL binds to the hypoxia inducible factor 1 (HIF-1) and triggers ubiquitin-mediated degradation via the proteasome of one of the HIF-1 subunits, HIF-1α and HIF-2α (two isoforms of the same subunit) [46, 47]. Accumulation of both HIF-1α and HIF-2α can drive a hypoxia-induced genetic program upon translocation from the cytoplasm into the nucleus via hypoxia-response elements (HREs) [48] (**Fig. 5**). Examples of HIF-regulated genetic programs are VEGF signaling, MYCB signaling, TGF-β signaling and Wnt/β-catenin signaling which all facilitate the growth and spread of the tumor [49]. Disruption of the HIF-1 pathway through VHL also

impacts Paired box (PAX) proteins in ccRCC. For example, PAX2 is highly active during kidney development and important for nephrogenesis. PAX2 is silenced post nephron maturation, but is commonly re-activated in ccRCC and up-regulated together with HIF-1α and HIF-2α, and drives tumor differentiation [50].

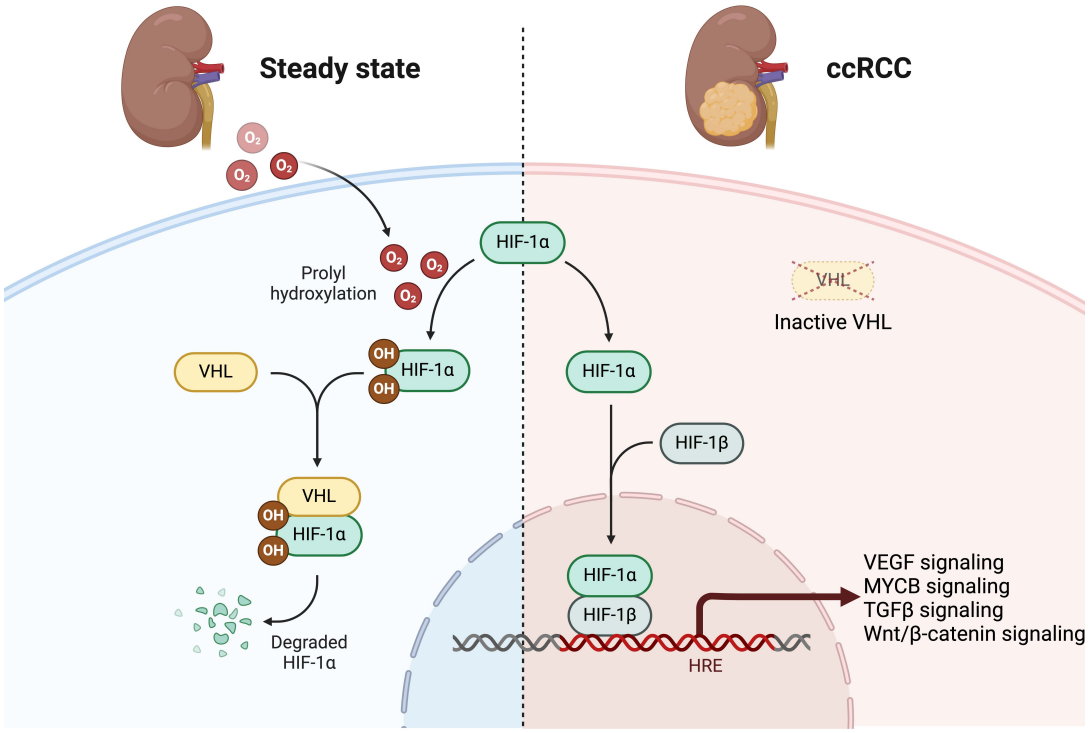


Figure 5. HIF-1 dependent signaling in the oncogenesis of ccRCC. Schematic depicting VHL-mediated, proteasome-dependent degradation of HIF-1α in normal kidney epithelial cells at steady state (left). VHL inactivation leads to a hypoxic, HIF-1 driven gene activation program in ccRCC (right). Schematic created with BioRender.com.

Despite their many complementary roles, HIF-1α and HIF-2α also have non-redundant functions at steady-state and tumorigenesis [47, 51, 52]. For example, it has recently been shown that co-option of PAX8 in a HIF-2α, and not HIF-1α, dependent manner drives a strong oncogenic signaling program in ccRCC [53]. Specifically, HIF-2α is recruited to transcriptional enhancers bound by PAX8, which also includes the pro-tumorigenic cyclin D1 (CCND1). Despite this clearly identified mechanism through which VHL promotes ccRCC, loss of VHL is not solely responsible for tumorigenesis. Multiple additional co-drivers have been identified, such as mutations in PBRM1, SETD2, KDM5C, and BAP1 [41, 54].

Despite the high mortality rates of ccRCC as compared to other RCCs, if discovered early ccRCC is relatively easy to treat compared to many other malignancies, mainly due to the ease

of applying curative surgery [55]. Depending on the size of the tumor and intra-kidney localization, either partial nephrectomy or radical nephrectomy, in which part of or the entire kidney is removed, is performed [56]. For patients with high risk of recurrence post nephrectomy, it has been recently discovered that adjuvant Pembrolizumab (an anti-PD-1 antibody) significantly increased the disease-free survival time, as well as the overall survival time, after surgery [57-59].

Other than surgery, the most common treatments for ccRCC are the use of tyrosine-kinase inhibitors (TKIs) or immunotherapy, either on their own or in combination with one another [60]. Examples of common TKIs for ccRCC are Sunitinib, Sorafenib, Cabozantinib and Pazopanib [61, 62]. All of these TKIs inhibit multiple tyrosine kinase receptors such as VEGFR and PDGF-R, with the goal of preventing angiogenesis and tumor cell proliferation [63-65]. Despite the promise of these therapies, roughly a quarter of patients do not respond to TKIs as a first-line therapy and, amongst the responders, patients rapidly develop resistance, leading to subsequent tumor growth and metastasis [66, 67]. The widespread development of drug resistance, or unresponsiveness to first-line treatments, has led to ccRCC tumors frequently being treated with second-line treatments. These treatments are either other TKIs, or immunotherapies which, in some cases, have been shown to have synergistic effects with TKIs. Recently, it is also getting increasingly common to use immunotherapies as first-line treatment followed by TKIs or other immunotherapies as second-line treatments [68-71]. Immunotherapies that are commonly used in ccRCC are PD-1/PD-L1 inhibitors such as Pembrolizumab [69, 72], Avelumab [73] or Nivolumab [71] and CTLA-4 inhibitors such as Ipilimumab [71, 74]. ccRCC tumors display high immune infiltration and the introduction of immunotherapies into the treatment-landscape of ccRCC initially brought on very promising results [75]. However, just like for TKIs, resistance to immunotherapies is a problem in ccRCC [76, 77].

Engineered cell therapies, such as chimeric antigen receptor T (CAR-T) cells, are gaining increased traction in cancer treatment [78]. In ccRCC, CAR-T cells targeting CD70 are currently being tested. In a small clinical trial with 16 patients, a complete response was achieved in 1 patient, whereas the other 15 exhibited mixed responses [79]. It is still too early to conclude how well cellular therapies will perform in ccRCC, but multiple clinical trials are currently ongoing with results coming soon [80].

One of the main reasons why treatments of ccRCC have had mixed results is the fact that patients frequently present in the clinic with already progressed disease. As mentioned, if detected early ccRCC can be readily treated, whereas later stages, and especially metastatic

disease, have a very poor prognosis with an average 5-year survival of only 12% [29, 81]. Therefore, there is an unmet need for better early-stage diagnosis methods and public health screening programs in order to detect the disease as early as possible for favorable treatment outcomes, as well as to identify prognostic markers to stratify patients into beneficial treatment groups considering the currently available treatment combinations.

1.3.2 Clear cell renal cell carcinoma diagnosis and prognosis

Diagnosis of ccRCC is a challenge in the oncology field. ccRCC tends to be asymptomatic for early stages of the disease resulting in patients presenting at clinics with already advanced disease. Obvious signs of advanced disease that patients sometime report prior to seeing a physician are e.g. palpable lumps on the backside adjacent to the kidneys and hematuria [82]. These signs are in and of themselves far from sufficient for a reliable diagnosis of kidney cancer, and do not indicate any molecular subtype the malignancy belongs to. Once a tumor is suspected, the patient usually undergoes imaging in which a renal mass is confirmed [83, 84], followed by histopathological assessment of the tumor subtype.

The gold standard for discerning between different subtypes of RCC is histopathological assessment of the tumor where nuclear morphology and cytological features are interrogated [85]. More sophisticated histological classification methods are being developed which implement machine learning of histopathology sections as well as gene expression profiling to increase the accuracy and speed with which different subtypes can be classified [86]. Despite high accuracy and increased turnaround times, histopathology remains time-consuming and is still invasive to the patient. Furthermore, improving the subtype determination of RCCs does not counteract the fact that tumors often are initially found at late stages. Therefore, we need better non-invasive, longitudinally accessible biomarkers for early-stage diagnosis and disease monitoring [82].

Urine would be an ideal source of diagnostic biomarkers due to the ease of sampling, non-invasive repeated sampling, and the kidneys' intrinsic role in filtering urine. Disruption of the steady-state GFR rate is commonly seen in ccRCC patients [87, 88]. Disturbed filtration, and tissue damage to the nephron caused by the tumor, could lead to dysregulated solute levels in the urine which can be probed for diagnostic purposes. Furthermore, ccRCC is largely found in the senior population, age 60 and upwards, and rarely younger individuals [82]. This aging population is often part of continuous health check-ups in which urine is sampled, making public health screening programs in the most at-risk population feasible.

For prognostic determination of ccRCC tumors, more tools are available. For example, MRI-based radiomics have shown promise in predicting ccRCC size, grade and necrosis degree before any surgery to stratify patients into prognostic groups [89]. Radiogenomics, the linkage of genetic phenotypes with imaging phenotypes [90], has been shown to be able to give insights into the genetic makeup of individual tumors non-invasively, and also provide prognostic predictions [91]. Other than imaging, molecular techniques have shown that e.g. the ClearCode34 gene signature, comprising the expression levels of 34 genes, can stratify patients into prognostic groups [92, 93]. Lastly, protein tissue arrays have been used, and integrated with clinical parameters such as grade and stage, to predict the survival of patients with ccRCC [94]. The drawback with these methods, despite their impressive results, is that they rely on imaging or molecular profiling of the solid tumor itself. Ideally, in a clinical setting, prognostic biomarkers for ccRCC would be easily assayable markers found directly in serum or urine that can be measured using standardized diagnostic methods such as ELISA. Therefore, in addition to needing readily assayable diagnostic biomarkers for ccRCC in the clinic, we also need better prognostic biomarkers.

1.5 Urine as a source of biomarkers

Urine is an excellent fluid for routine access of prognostic or diagnostic material. Urine is easily accessible and can be continuously sampled by subjects themselves and the collection procedure is minimally invasive [95-97]. The first diagnostic cancer biomarker was actually derived from urine in 1848, the Bence Jones protein, and has historically been used to diagnose multiple myeloma [98]. For certain urogenital cancers, the development and clinical use of urinary biomarkers is also underway.

For bladder cancer, FDA-approved protein-based methods have been available for disease monitoring and diagnostics for a while but do not perform as well as conventional ureteroscopic, cystoscopic and cytological evaluation of the urinary tract and urine. However, these are invasive methods associated with a disproportionally high cost from first diagnosis to death compared to other types of cancers [99]. New protein-based diagnostics have recently been developed and perform better than previous tests relying on singular proteins for diagnosis. For example, Oncuria™ has developed a multiplexed immunoassay to diagnose detect bladder cancer, assessing ten different proteins and achieved an overall area under receiver operating characteristic curve (AUROC) of 0.95 [100]. A receiver operating characteristic (ROC) analysis is a graphical way of assessing a model's performance, e.g. in

this case in being able to distinguish between two different states: healthy control or cancer. To construct a ROC curve, the true positive rate (TPR) is plotted against the false positive rate (FPR) at all thresholds. An ideal model would have a TPR of 1 and FPR of 0, and the AUROC would then be 1 [101].

Furthermore, progress has been made utilizing other sources than protein for diagnosis in liquid urine biopsies [102]. For example, the UroSEEK platform [103, 104] and the Urine Cancer Personalized Profiling by deep Sequencing (uCAPP-Seq) platform [105] have shown that mutated urinary tumor DNA (utDNA) detection can be used for the detection and surveillance of bladder cancer with high sensitivity and specificity, and the uCAPP-Seq platform reported AUROC values of 0.92 in patients with no a priori knowledge about the genetic makeup of the tumor [106]. Other urinary liquid biopsy methods for bladder cancer have been extensively reviewed elsewhere [106, 107].

Whereas the field of urinary liquid biopsies for bladder cancer has been extensively researched in the last years resulting in FDA approvals and clinical trials, renal cell carcinomas are far behind. The below section details the studies that have explored proteins as a way of diagnosing ccRCC. Differences in urine protein content has been found amongst ccRCC patients depending on disease prognosis [108, 109] and venous infiltration [110]. Furthermore, attempts have been made at identifying diagnostic and prognostic protein biomarkers in the urine of ccRCC patients [97, 111]. For example, the proteins EPS8L2 and CHMP2A have shown low performance in distinguishing between progressed and non-progressed clear cell renal cell carcinoma, with AUROC values of 0.76 and 0.7, respectively [111]. When attempting to find diagnostic biomarkers, one study found that combining the protein levels of PKHL1L1, C3, FABP4 and ANGPTL6 could distinguish between controls and ccRCC with a performance accuracy of AUROC = 0.835 [97]. Lastly, one study found that PLIN-2 is highly elevated in the urine of ccRCC patients and can be robustly detected with a paper-based bioplasmonic assay [112]. Further, when expanding the method to also encompass AQP-1, the performance was increased in a AUROC analysis to 0.990 in a small cohort [113]. However, this method included both ccRCC and pRCC patients and cannot distinguish between them and has not been developed further since.

Other than proteins, nucleotide-based readouts have also been investigated as a way to diagnose ccRCC in urinary liquid biopsies with varied results. For example, miR-30a-5p^{me} levels had a low AUROC of 0.6837 in distinguishing between ccRCC and controls but performed better in tissue samples [114]. Other miRNA approaches are reviewed elsewhere [115]. Further, a gene panel looking into the methylation status of *PCDH17* and *TCF21* achieved an AUROC of 0.656

in distinguishing between ccRCC and controls. The most promising results come from a study using cell-free methylated DNA immunoprecipitation and high-throughput sequencing (cfMeDIP–seq) in which sequencing of a panel of 300 differentially methylated regions could distinguish between controls and patients with an AUROC of 0.858, but with a much higher performance in plasma in a mixed RCC cohort [116]. The downside to the cfMeDIP–seq approach is that it requires extensive laboratory work with immunoprecipitation and next generation sequencing. The discussed urinary liquid biopsy diagnosis markers aresummarized in **Fig. 6**.

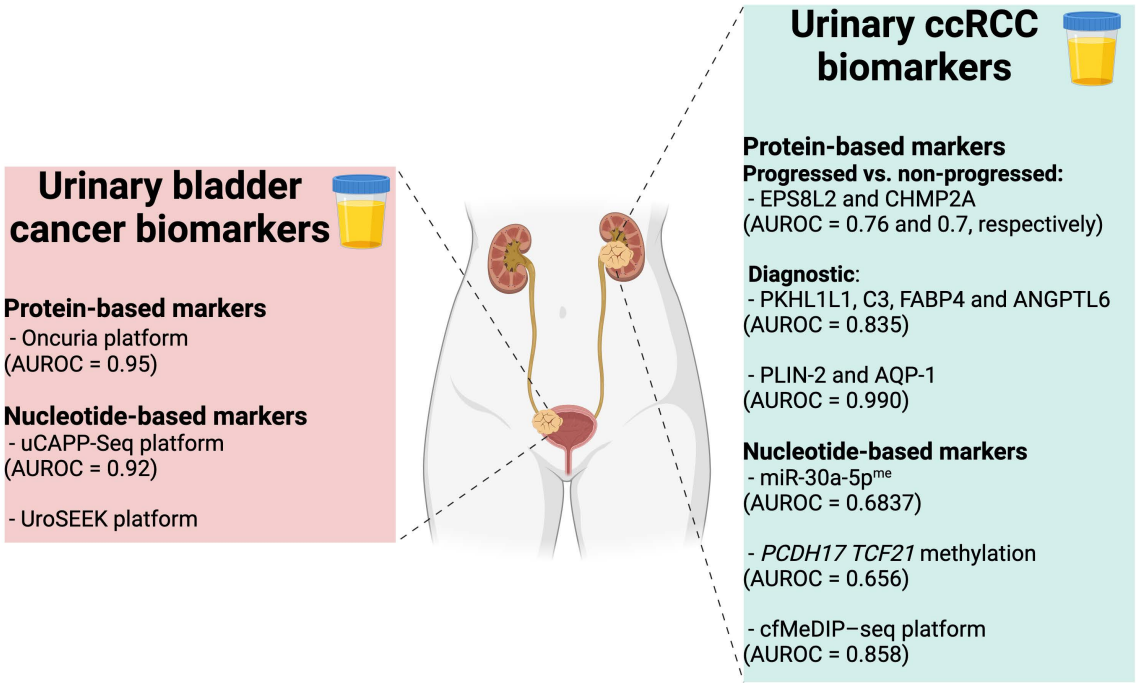


Figure 6. Urinary liquid biopsy biomarkers in bladder cancer and ccRCC. Summary of some protein and nucleotide-based markers currently under investigation in bladder cancer and ccRCC. Figure created with BioRender.com.

With the overall moderate performance of current urine-based biomarkers for ccRCC, further research efforts are needed to elucidate high-quality and readily assayable biomarkers.

1.6 Secreted biomarkers for ccRCC

ccRCC predominantly arises from proximal tubular epithelial cells within the proximal nephrons [117]. The proximal tubules are the main reabsorption units in the nephron and thus directly filter molecules between the blood and urine [13, 14, 118]. Therefore, we hypothesized that upregulated secreted markers directly from the tumor microenvironment could be present

at higher levels in the urine of ccRCC patients due to disrupted GFR and/or tissue damage to the proximal tubule resulting in leakage. Urine could thereby be explored as a diagnostic and prognostic tool for ccRCC.

To test the feasibility of using secreted biomarkers to detect or predict outcome of ccRCC, we first performed an *in silico* screen on The Cancer Genome Atlas (TCGA) [119] transcriptomics data from ccRCC (TCGA annotation: KIRC) (**Fig. 7A**). TCGA RNA sequencing data from tumor and normal tissue for TCGA-KIRC was downloaded through the GDC Data Portal [120]. Differential expression analysis between tumor and matched normal tissue was performed with DESeq2 (v1.22.2) [121]. Significantly upregulated or downregulated genes were defined as genes with ≥10 counts, and Benjamini-Hochberg adjusted Wald p values of $p < 0.05$. For secretion prediction, the Human Protein Atlas (HPA) [122, 123] secretome classification was used, in which secreted proteins are defined as having a signaling peptide and no transmembrane domain, resulting in a list of 2641 secreted transcripts [124].

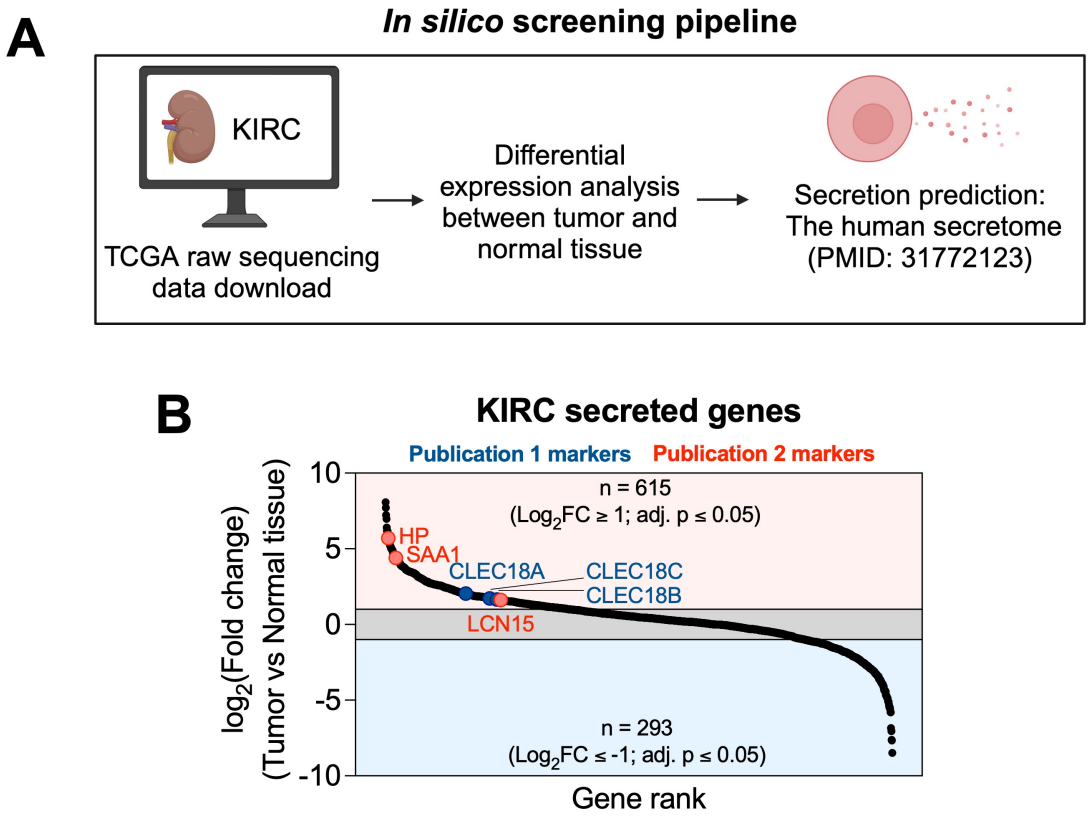


Figure 7. *In silico* screening on TCGA-KIRC data reveals putative secreted biomarkers. (A) Schematic of how the screen was performed. Downloaded raw RNA-seq data from the GDC data portal for ccRCC/KIRC was subjected to the DESeq2 differential expression pipeline and filtered for secreted hits using the human secretome. **(B)** Gene rank plot of significantly upregulated and downregulated genes encoding secreted proteins in ccRCC/KIRC. Genes that were studied further in **Publication 1** and **Publication 2** are highlighted.

Differential expression analysis was performed between tumor and normal tissue for KIRC and all non-secreted genes were removed. A total of 615 upregulated genes remained as potential diagnostic and/or prognostic biomarkers after excluding non-secreted genes and implementing a cut-off of $\log_2$ fold change > 1 (**Fig. 7B**). Markers of interest discussed further in **Publication 1** (CLEC18A, CLEC18B and CLEC18C) and **Publication 2** (SAA1, HP and LCN15) are highlighted.

1.7 C-type lectins in cancer

Three of the main markers which were studied further in this thesis in **Publication 1** is a family of C-type lectins, the CLEC18A family. Collectively, C-type lectins are glycan binding proteins with carbohydrate recognition domains (CRDs). The binding between the lectin and the carbohydrate is conventionally dependent on the presence of Ca^{2+} [125].

C-type lectins have predominantly been studied in the context of pathogen recognition. C-type lectins have for example been described as binding to various fungi, such as various species of the genera *Candida*, *Aspergillus* and *Pneumocytis* [126], as well as multiple different bacterial species [127, 128]. Furthermore, C-type lectins have been found to also bind viruses, for example it was recently found that some C-type lectins can bind to the Spike protein of SARS-CoV-2 [129].

In recent years, we have also started gathering a lot more knowledge regarding the role of C-type lectins in cancer. In a cancer setting, C-type lectins have been found to have multiple, and sometimes, opposing roles where they can either promote or inhibit tumor growth. For example, the C-type lectins Dectin 1, Mincle, MGL and DC-SIGN are all myeloid restricted C-type lectins and have all been found to repress the immune system, subsequently promoting tumor growth [130-132]. On the other hand, some C-type lectins have very clear roles in inhibiting tumor growth [133, 134]. One of the most well-studied examples of an anti-tumor C-type lectin is NKG2D which acts as a very potent cytotoxic receptor on NK cells [135].

The knowledge regarding C-type lectins and their role in cancer is growing, especially from an immunological standpoint. However, whether or not they can, and will, be implemented into clinical practice therapeutically, diagnostically or prognostically remains to be seen.

2. Aims and results of the PhD project

In my PhD project, I set out to establish new prognostic and diagnostic biomarkers for clear cell renal cell carcinoma from urine. As discussed, urine is an attractive source of biomarkers for renal malignancies due to hematuria, proteinuria, low invasiveness, and ease of sampling. We hypothesized that upregulated markers in the ccRCC microenvironment which are secreted would be present at higher levels in the urine and could serve as putative prognostic and/or diagnostic biomarkers. Technically, a marker does not have to be upregulated in a disease to serve as a prognostic marker. However, due to the overall low protein concentrations in urine at steady state we opted for upregulated markers compared to healthy controls to make sure that the markers would be able to be reliably measured above the detection threshold. Further, if the prognostic marker is also upregulated it could serve as both a prognostic and diagnostic biomarker.

As a proof of concept, we performed an *in silico* screen on The Cancer Genome Atlas (TCGA) data for ccRCC and found more than 600 transcripts which are upregulated and secreted in ccRCC tumors compared to healthy kidney tissue (**Fig. 7**).

The *in silico* screen revealed three interesting transcripts that previously have not been linked to ccRCC: the CLEC18 gene family. The CLEC18 gene family of C-type lectins consists of three paralogs in human (CLEC18A, CLEC18B and CLEC18C). All three of the CLEC18 paralogs were upregulated in ccRCC compared to normal tissue and are all predicted to be secreted. The expression of the CLEC18 family was shown to be restricted to the proximal tubules of the kidney and the medial habenula of the brain. C-type lectins predominantly interact with carbohydrates through their carbohydrate recognition domain (CRD). Through phylogenetic analysis it was shown that CLEC18 proteins have a unique CRD sequence and experimentally they were shown to likely interact with glycosaminoglycans of proteoglycans, which is a unique ligand for C-type lectins. Lastly, we were able to experimentally show that high expression of the CLEC18 family positively correlates with survival of ccRCC patients and could be used as a prognostic marker. Furthermore, loss of function mouse studies confirmed these findings. Combined, these results make up the contents of **Publication 1**.

Following the proof-of-concept *in silico* screen, mass spectrometry-based proteomics was performed on urine samples from ccRCC patients, nccRCC patients and healthy controls as a validation study. Proteomics validated three proteins with high confidence: Serum amyloid A1 (SAA1), Haptoglobin (HP) and Lipocalin 15 (LCN15). Furthermore, a parallel reaction monitoring mass spectrometry (PRM-MS) method was established for rapid and sensitive

readout of SAA1, HP and LCN15 in urine samples. Our PRM-MS method performed very well in distinguishing between ccRCC and controls, and also between ccRCC and ncRCC patients. Lastly, proteomics, metabolomics and lipidomics of the urine also revealed an overall dysregulated metabolism in ccRCC patients compared to controls marked by increased lipid transport. Combined, these results are reported in **Publication 2**.

The results from both **Publication 1** and **Publication 2** are discussed in more depth in the Discussion section below.

3. Publications

My PhD work has resulted in two first author publications which are available as preprints on bioRxiv [43, 136]. Both papers have been submitted for peer-reviewed publication, and my contribution to both manuscripts is clearly stated herein and in the preprints. Proof of submission for both publications is included at the end of this thesis.

3.1 Publication 1: CLEC18A interacts with sulfated GAGs and controls clear cell renal cell carcinoma progression

Authors:

Gustav Jonsson[1,2,3,*], Maura Hofmann[1,3], Stefan Mereiter[1,3], Tiago Oliveira[1,3], Irma Sakic[1], David Hoffmann[1], Maria Novatchkova[1,4], Alexander Schleiffer[1,4], Josef M. Penninger[1,3,5,6,*]

Affiliations:

1. Institute of Molecular Biotechnology of the Austrian Academy of Sciences, IMBA, Dr. Bohr-Gasse 3, 1030 Vienna, Austria
2. Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, 1030, Vienna, Austria
3. Eric Kandel Institute, Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
4. Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, 1030 Vienna, Austria
5. Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Canada
6. Helmholtz Centre for Infection Research, Braunschweig, Germany

* Correspondence to: Josef M. Penninger (josef.penninger@helmholtz-hzi.de) and Gustav Jonsson (gustav.jonsson@imba.oeaw.ac.at)

Author contributions:

G.J. and J.M.P conceived the study. G.J. performed and designed all experiments with input and help from all co-authors as follows: M.H. with cell line generation, mouse experiments and

AlphaFold2 Multimer screens, S.M. and T.O. with CLEC18 ligand identification, I.S. with mouse experiments, D.H. with recombinant CLEC18A-Fc generation, M.N. with TCGA analysis and A.S. with phylogeny analysis.

CLEC18A interacts with sulfated GAGs and controls clear cell renal cell carcinoma progression

Gustav Jonsson[1,2,3,*], Maura Hofmann[1,3], Stefan Mereiter[1,3], Lauren Hartley-Tassell[4], Irma Sakic[1], Tiago Oliveira[1,3], David Hoffmann[1], Maria Novatchkova[1], Alexander Schleiffer[1], Josef M. Penninger[1,3,5,6,*]

1. Institute of Molecular Biotechnology of the Austrian Academy of Sciences, IMBA, Dr. Bohr-Gasse 3, 1030 Vienna, Austria
2. Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, 1030, Vienna, Austria
3. Eric Kandel Institute, Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
4. Institute for Glycomics, Griffith University, Southport, QLD, Australia
5. Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Canada
6. Helmholtz Centre for Infection Research, Braunschweig, Germany

* Correspondence to: Josef M. Penninger (josef.penninger@helmholtz-hzi.de) and Gustav Jonsson (gustav.jonsson@imba.oeaw.ac.at)

Abstract

C-type lectins are a large family of proteins with essential functions in both health and disease. In cancer, some C-type lectins have been found to both promote and inhibit tumor growth, but many of the C-type lectins still remain uncharacterised in a tumor context. Therefore, there is growing interst in further elucidating the mechanisms with which C-type lectins control tumor growth. Here, we report a key role of the CLEC18 family of C-type lectins in the progression of clear cell renal cell carcinoma (ccRCC). The CLEC18 family is conserved across the entire Chordata phylum with recent gene duplication events in humans. We found that CLEC18A is exclusively expressed in the proximal tubule of the kidney and the medial habenula of the brain. We further identified sulfated glycosaminoglycans (GAGs) of proteoglycans as the main CLEC18A ligand, making them unique among C-type lectins. In ccRCC patients, high expression of the *CLEC18* family lectins in the tumor are associated with improved survival. In mouse models of ccRCC, deletion of the mouse ortholog *Clec18a* resulted in enhanced tumor growth. Our results establishes CLEC18A as a novel and critical regulators of ccRCC tumor growth and highlights the potential benefit of modulating *CLEC18* expression in the renal tumor microenvironment.

Keywords: C-type lectin, glycosaminoglycan, proteoglycan, proximal tubule, clear cell renal cell carcinoma

Introduction

C-type lectins are a vast superfamily of proteins found across the entire vertebrate lineage with important functions in physiology and disease. The C-type lectins are defined by containing at least one C-type lectin domain (CTLD), functioning as their carbohydrate recognition domain (CRD). Classically, C-type lectins have been studied in the context of both innate and adaptive immunity towards microbial pathogens, but C-type lectins also play key roles in autoimmunity and cancer [1-3]. Overall, C-type lectins show varied evolutionary conservation across vertebrates. Some C-type lectins are specific to human, and others only have weak orthology to commonly used model species making them difficult to study in a disease context [4-6].

However, in the last decades, key insights have been generated into how C-type lectins are involved in cancer. Some C-type lectins, such as Dectin 1, Mincle, MGL and DC-SIGN, have all been found to suppress immunity towards tumor cells through their expression in the myeloid compartment [7-9]. On the other hand, C-type lectins have also been found to be involved in eradicating tumor cells. For example, the C-type lectin NKG2D recognises polymorphic MHC class I related stress induced ligands and serves as a key activating receptor for NK cells [10]. Further, myeloid restricted C-type lectins have been shown to be involved in the clearance of liver metastasis [11] and suppression of hepatocarcinogenesis [12].

Among C-type lectins the secreted CLEC18 family remains poorly characterised. The human *CLEC18* locus, consisting of the paralogs *CLEC18A, CLEC18B* and *CLEC18C*, is located on chromosome 16q22 and encodes for a family of secreted proteins [13]. In viral infections, *CLEC18* has been shown to be downregulated in Hepatitis B infections while upregulated during Hepatitis C infections [14, 15]. Furthermore, CLEC18 associates with TLR3 and thereby stimulates the production of type I and type III interferons following H5N1 influenza A virus exposure [16] and transgenic expression of CLEC18A in *Aedes aegypti* mosquitos significantly reduced dengue virus infectivity [17]. Other than these reports, very little is known about the role of CLEC18 in normal physiology and other diseases, such as cancer. Moreover, the exact carbohydrate ligands for CLEC18 family members remains elusive.

In the present study, we provide a deep phylogenetic analysis of the *CLEC18* family and its orthologs throughout the Chordata phylum. Expression of CLEC18A was mapped to the medial habenula of the brain and the proximal tubule of the kidney. We also resolved the first high-confidence CLEC18A ligands, sulfated glycosaminoglycans (GAGs) on a collection of

proteoglycans. Interestingly, *CLEC18* expression was associated with favorable cancer prognosis in ccRCC. Lastly, loss of function studies in mice showed that deletion of *Clec18a* in renal cancer cells resulted in a significant increase in the tumor burden of mice.

Results

CLEC18 is a family of conserved gene paralogs restricted to the kidney and brain

To explore evolutionary conservation, we identified CLEC18 orthologs across vertebrates. We found that that human CLEC18A has homologs and paralogs across the Chordata phylum with the earliest identifiable ancestor being lampreys. Unlike some other C-type lectins which lack direct orthology between species [4], CLEC18A instead has different number of paralogs in several species. Lower vertebrates and most mammals only have one copy of CLEC18, whereas some higher species have two or three paralogs, e.g. humans (CLEC18A, CLEC18B, and CLEC18C) (**Supplementary Fig. 1A**), indicating recent gene duplication events. Furthermore, the three human CLEC18 paralogs are almost identical (**Fig. 1A**, **Supplementary Fig. 1B**) and across the entire Chordata phylum the CLEC18 proteins are highly conserved and ordered (**Fig. 1B-D**), when present. For example, no CLEC18 homologs are found in the common model organisms zebrafish and xenopus.

Next, we determined the expression pattern of *Clec18a* mRNA. RT-qPCR of mouse organ revealed highly restricted expression to the brain and kidney, with the kidney exhibiting the highest expression (**Fig. 1E**, **Supplementary Fig. 1C**). RNA fluorescence *in situ* hybridization of the brain revealed that *Clec18a* is almost exclusively expressed in the medial habenula (**Fig. 1F**). Re-analysis of the kidney cell atlas [18] of fetal kidneys revealed that human *CLEC18A* is predominantly expressed in various structures of the fetal nephron (**Supplementary Fig. 2A**). A more detailed assessment of human fetal nephrons revealed that during development *CLEC18A* is most commonly found in the distal renal vesicle, proximal tubules and S-shaped bodies (**Fig. 1G**), the latter of which gives rise to the proximal and distal tubules, and the loop of Henle. In the adult kidney, *CLEC18A* was almost exclusively found in the proximal tubule (**Fig. 1H,** full clustering shown in **Supplementary Fig. 2B**). These data show that CLEC18A is a highly conserved and ordered C-type lectin protein across the Chordata phylum, predominantly found in the medial habenula and proximal tubule of the kidney.

CLEC18A interacts with glycosaminoglycans (GAGs) on proteoglycans

We next sought to elucidate the ligand of CLEC18 proteins. Phylogenetic analysis and principal component analysis on a multiple sequence alignment of the CRDs of canonical C-type lectins (**Fig. 2A, Supplementary Fig. 3A**) revealed that the CLEC18 paralog family forms its own cluster (**Fig. 2B**), indicating that they have unique CRD sequences and that they might interact with unique ligands. A previous study reported glycoarray analysis with a CLEC18A construct using an array with 611 common N-linked and O-linked glycostructures, but failed to elucidate significant interaction partners [13]. However, despite not reaching significance, the strongest hit was GlcAβ1-6Galβ (**Supplementary Fig. 3B**). Glucuronic acid (GlcA) is primarily used as a building block in glycosaminoglycans (GAGs) and appears in all glycosaminoglycans of proteoglycans except for keratan sulfate. Furthermore, GlcA linked to Galactose (Gal) appears at the start of heparan sulfate and chondroitin sulfate, albeit not with a β1-6 linkage [19]. Assuming that the correct ligand was not included in the original array, these results suggested to us that a proteoglycan GAG could be the ligand.

To determine if proteoglycans indeed are the ligands of CLEC18, we expressed a fusion between the murine CRD of CLEC18A and an antibody Fc region (CLEC18A-Fc) (**Supplementary Fig. 4A**). The binding of CLEC18A-Fc was then screened against a large panel of cancer cell lines. The strongest binder (PymT, a murine mammary cancer cell line) and one non-binder (Ac711, also a murine mammary cancer cell line) were chosen for downstream analysis (**Supplementary Fig. 4A-B**). Using these two cell lines, multiple putative CLEC18A ligands were identified in a lectin-pulldown approach (**Fig. 2C, Supplementary Table 1**), many of which were Proteoglycans (e.g. Biglycan (*Bgn*), Vestican (*Vcan*), Syndecan 1 (*Sdc1*) and Syndecan 2 (*Sdc2*)). All putative ligands were filtered for proteins that are either secreted or membrane bound since CLEC18A is predicted to be localized to the extracellular matrix, hypothesizing that the exterior of the cell or extracellular space would be the site of interaction. Amongst this annotated list of ligand candidates, multiple proteoglycans and proteoglycan related proteins were found (**Fig. 2D**).

To investigate whether or not the identified interactions in the pulldown actually are GAGs on proteoglycans, or interactions with the protein backbone, we performed an AlphaFold2 Multimer [20-22] screen with human CLEC18A, CLEC18B and CLEC18C against the amino acid sequences of the secreted and membrane-bound hits identified in the pulldown. In this approach we did not identify any highly scoring protein-protein interactions (PPIs), indicating

that the interactions are between the CRD of CLEC18 and the GAG of the proteoglycan (**Fig. 3A**). Furthermore, since CLEC18A is expressed in the proximal tubule, we also performed an AlphaFold Multimer screen with CLEC18A, CLEC18B and CLEC18C, but this time against a kidney specific library of proteins consisting of 379 unique amino acid sequences. In this approach, there were also no identified high-confidence PPIs (**Fig. 3B**). To confirm the GAG specificity of CLEC18A, we used our CLEC18A-Fc fusion protein to probe a glycoarray with GAGs and other glyco-structures (**Fig. 3C**). The glycoarray contained a total of 254 glyco-spots. After removing unspecific binding of the Fc-control, 242 structures remained and out of those 29 glycostructures were identified as glyco-ligands of CLEC18A (**Fig. 3D, Supplementary Table 2**). CLEC18A bound to a variety of heparin, heparan sulfate, and chondroitin sulfate (both low and high molecular weight) GAGs. CLEC18A did not interact with any hyaluronan and hyaluronan fragments which are un-sulfated, indicating the need for disaccharide sulfation for CLEC18A docking (**Fig. 3E**). Of note, CLEC14A is the only other C-type lectin which has been reported to interact with proteoglycan GAG chains [23]. Taken together, these data show that CLEC18A interacts with sulfated GAGs on proteoglycans.

Association of CLEC18 expression with kidney and brain cancer patient survival

CLEC18A has been found to be regulated by, or involved in, some viral infections, such as Hepatitis C and Dengue fever [15-17]. Whereas many C-type lectins play a role in host-pathogen interactions, C-type lectins are also commonly involved in cancer [24, 25]. To assess whether *CLEC18A*, *CLEC18B* or *CLEC18C* might have a role in cancer, we first analyzed patient survival data for all cancer types in The Cancer Genome Atlas (TCGA). Interestingly, we found that expression of all three human *CLEC18A*, *CLEC18B* and *CLEC18C* paralogs is correlated with the survival of patients with clear cell renal carcinoma (ccRCC, TCGA abbreviation: KIRC) (**Fig. 4A**). Of note, out of the three kidney cancer types present in the TCGA; KIRC, papillary renal cell carcinoma (KIRP) and chromophobe renal cell carcinoma (KICH), the expression of *CLEC18* gene cluster only correlated with the survival of KIRC patients **(Fig. 4A).** The individual survival curves for these observations are shown in **Fig. 4B-D** for *CLEC18A* and **Supplementary Fig. 5A-F** for *CLEC18B* and *CLEC18C*. Moreover, mRNA expression of the three CLEC18 family members was upregulated in the tumors of KIRC patients but downregulated in KIRP and KICH **(Fig. 4E**). Lastly, *CLEC18* expression was inversely correlated with the survival of patients with Low grade gliomas (LGG) (**Fig. 4A**). These data indicate that, in line with the very restricted expression profiles of the *CLEC18* family members in the brain and kidney, *CLEC18* expression levels are associated with the

survival of patients with low grade glioma and clear cell renal cell carcinomas. These data also show that, for KIRC/ccRCC specifically, that the *CLEC18* gene family is upregulated in the tumors and higher expression correlates with better survival for the patients.

Loss of CLEC18A accelerates kidney cancer cell growth in mice

To further test a potential role of *CLEC18A* in ccRCC, we utilised murine cell lines which only have one CLEC18 paralog, CLEC18A. A murine ccRCC cell line called RAG, which exhibited a high baseline expression of *Clec18a*, was used to generate *Clec18a* knockouts (*Clec18a*$^{-/-}$) using CRISPR-Cas9 (**Fig. 5A**). As a control, we overexpressed *Clec18a* (*Clec18a*^{OE}) in an unrelated murine mammary cancer cell line, E0771, which did not have any baseline expression of *Clec18a* (**Fig. 5B**). Deletion or introduction of *Clec18a* had no impact on the *in vitro* growth pattern of either cell line (**Fig. 5C-D**).

Based on the TCGA data where we found that high *CLEC18A* expression favors survival, we hypothesized that deletion of *Clec18a* in the ccRCC cell line may promote tumor growth. Indeed, genetic inactivation of *Clec18a* promoted tumor growth of RAG ccRCC cancer cells following subcutaneous injection into mice (**Fig. 5E**). Of note, these experiments were performed in *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice which lack T, B and NK cells since the RAG cell line did not grow in immunocompetent mice. Thus, these data indicate that the murine Clec18a hinders tumor progression in a lymphocyte-independent manner. We did find a significantly increased tumor infiltration of Ly6G^{+} neutrophils in *Clec18a*$^{-/-}$ ccRCC tumors (**Fig. 5F**) but no apparent difference in infiltrating F4/80^{+} macrophages (**Fig. 5G**). Overexpression of *Clec18a* had no impact on the tumor progression in the murine breast cancer model in *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice (**Fig. 5H**). In conclusion, our study has shown that the CLEC18 family of proteins are highly conserved. Furthermore, *CLEC18A* is predominantely expressed in the proximal tubule of the kidney and interacts with sulfated GAGs on proteoglycans. Lastly, *CLEC18* is upregulated in the tumor microenvironment of ccRCC and the expression level is positively correlated with survival in both mice and humans (**Fig. 5I**)

Discussion

C-type lectins are known to play essential roles in cancer ranging from recognition of damaged cells and antigen uptake [26, 27], to elimination of cancer cells through MHC I associated stress signals [10]. In some cases, C-type lectins can also promote tumor growth [9]. The interest for pharmacologically targeting C-type lectins in the clinic for cancer treatments is steadily

increasing, but remains a challenging goal due to their complex and numerous mechanisms in malignancies [28]. Therefore, there is an increased need to further our understanding of C-type lectin biology in health and cancer.

In this study we identify a previously poorly characterised C-type lectin family, the CLEC18 family, as novel regulators of ccRCC progression. The CLEC18 gene family has varying numbers of paralogs across the Chordata phylum and higher numbers of paralogs in evolutionary proximity to humans, indicating recent gene duplications and/or other mutational events. Although quite rare across the entire genome, other C-type lectins also display recent evolutionary events. Furthermore, some C-type lectins lack clear orthologs between species and, just like CLEC18A, show species-specific duplications. Prominent examples of copy number differences are the DC-SIGN/SIGN [5, 29] and collectin [6] proteins. Interestingly, for both DC-SIGN/SIGN and collectin, the mouse has more paralogs than the human. The inverse relationship is observed for the CLEC18 gene family which has three paralogs in humans and only one in the mouse. Thus, generating and using murine systems, such as full body knockout or conditional knockout mice, in future studies will provide better models to probe CLEC18 function in physiology and disease, compared to human systems in which three genetic targets have to be manipulated. Adding to the genetic complexity of the CLEC18 family, multiple paralogous sequence variants and polymorphic variants have recently been identified across the three human paralogs including a novel segmental duplication in *CLEC18A* [30]. However, despite the genetic variations in human *CLEC18* genes, the gene family is highly conserved across the Chordata phylum with the CRD segment presenting the highest degree of conservation. The high conservation in the CRD indicates that the ligands of CLEC18 are also conserved across species. The high conservation and maintenance of several CLEC18 copies in multiple species would indicate that CLEC18 is essential, however we failed to find CLEC18 orthologs in some aquatic model organisms such as zebrafish and xenopus. The reason as to why some species lack CLEC18 remains to be determined.

Compared to other canonical C-type lectins, CLEC18 have a unique CRD amino acid sequence suggesting that they bind to unique ligands. Our data now show that CLEC18 predominantely interacts with the GAG chain of proteoglycans, namely a variety of heparin, heparan sulfate, and chondroitin sulfate GAGs. So far, CLEC14A has been shown to be the only other C-type lectin that interacts with proteoglycans [23]. Whereas many studies have linked various proteoglycans and their GAGs to a wide range of kidney diseases [31, 32], further studies are

needed to determine the exact role of the CLEC18-GAG axis in the context of health and disease of the kidney.

In cancer, we uncovered that CLEC18A is involved in the progression of ccRCC tumors. *CLEC18* genes are upregulated in ccRCC tumors and high CLEC18 expression correlates with improved survival of patients. When deleting *Clec18a* from murine ccRCC cells, these *Clec18a*$^{-/-}$ mutant kidney cancer cells grew faster and presented with more neutrophil infiltration. The exact molecular mechanisms by which the secreted CLEC18 lectins and their ligands affect kidney cancer growth remain unknown. However, it is known that some of the proteoglycans and their assocaited GAGs (that we have identified as ligands of CLEC18) promote cancer. Indeed, Versican, Biglycan, Syndecans and Glypicans have all been found to increase the migratory and metastatic capabilities of different cancers through various mechanisms [33-36]. Knowing that proteoglycans can contribute to the invasiveness and metastatic spread of cancer, and that metastatic ccRCC is highly lethal [37], it is possible that CLEC18 in the tumor microenvironment reduces tumor growth and possibly metastatic spread through its interaction with various GAGs. Future studies are needed to elucidate how the CLEC18-proteoglycan axis mechanistically contributes to the growth of tumor cells, and how this pathway could be pharmacologically explored as a novel treatment for kidney cancers.

Materials and methods

Phylogenetic analysis

To construct a phylogenetic tree of the CLEC18 family, orthologs of the protein were collected in individual blast searches starting with human representatives in both the NCBI protein database and UniProt reference proteomes [38-40]. An E-value threshold of $< 1x10^{-30}$was applied to identify significant matches. A selection of 37 hits covering a wide taxonomic range were then aligned using mafft (-linsi) [41], and columns with less than 20 sequences were removed (regions with long gaps). A maximum likelihood phylogenetic tree was calculated with IQ-TREE 2 v.2.2.2.6 [42], with standard model selection using ModelFinder [43]. We obtained branch supports with the ultrafast bootstrap (UFBoot2) [44] and the SH-aLRT test [45]. The tree was visualized in iTOL v6 [46] and rooted by the lamprey sequences. Branches that are supported by SH-aLRT $\geq$ 80% and UFboot2 $\geq$ 95% are indicated by a grey dot. Branch lengths represent the inferred number of amino acid substitutions per site, and branch labels are composed of gene name (if available), genus, species, and accession number. Conservation scores were calculated with Jalview using previously generated Multiple Sequence Alignments

(MSA) of all found CLEC18 proteins compared to human CLEC18A. Via Jalview calculated conservation scores are based on multiple sequence alignment analysis using the AMAS method [47]. Disorder probability of the human CLEC18A protein was calculated using the IUPred3 online tool [48]. Additionally, sequences of all known human C-type lectin (CLEC) CRDs were extracted from the UniProt database, aligned using mafft (-linsi) and analyzed using Jalview. The neighbor-joining tree based on observed divergence was calculated in seaview [49] and visualized with iTOL.

The 3D structure model of CLEC18A was derived from the AlphaFold Protein Structure Database [21, 50] and visualized with ChimeraX [51]. The per residue sequence conservation score was calculated with AAcon v. 1.1. (KARLIN method, results normalized with values between 0 and 1) [52].

RT-qPCR

Murine tissues were extracted and homogenized, and RNA isolated using an RNeasy mini kit (Qiagen). Total RNA was reverse-transcribed with the iScript cDNA Synthesis Kit (Bio-Rad). Real-time quantitative PCR was performed for murine *Clec18a* and *Gapdh* with GoTaq qPCR master mix (Promega) on the CFX384 system (Bio-Rad). Primer sequences: *Clec18a* forward: GCA GAC ACC TAC TAT GGA GCC A, *Clec18a* reverse: CAC TGT CAG TCA CCT CGT TGG T, *Gapdh* forward: GTC GGT GTG AAC GGA TTT GG, *Gapdh* reverse: GAC TCC ACG ACA TAC TCA GC.

scRNA-sequencing re-analysis of published datasets

Processed scRNA-seq data from the the kidney cell atlas [18] was downloaded from the CZ CELLxGENE platform [53, 54] as .rds files and the Seurat objects were imported into Rstudio. All re-analysis and visualization of the kidney cell atlas was performed with Seurat v5 [55].

CLEC18A-Fc fusion construct cloning, expression and purification

The CLEC18A-Fc fusion construct was cloned into the pCAGG_00_ccb plasmid as previously described [56] generating a construct with an inducible murine CLEC18A CRD fused to a murine IgG2a-Fc domain connected with a $(GGGS)_3$ linker. Before the construct is a IL2 secretion signal allowing purification of the CLEC18A-Fc fusion from culture supernatant. The CLEC18A-fusion was expressed by transfecting the plasmid into Freestyle™ 293-F cells. Cells were seeded at $0.7x10^6$ cells/ml the day before expression and maintained after transfection at

120 rpm, 37°C, 8% CO_2. The transfection was performed accordingly: 2 µl polyethylenimine (PEI) 25K (1 mg/ml; Polysciences) was mixed with each µg of plasmid in Opti-MEM medium (Thermo Fisher Scientific) and added to the cell suspension. 24h after transfection, the medium was topped up with EX-CELL 293 Serum-Free Medium (Sigma-Aldrich) to 20% and the culture was grown for 120h.

After expression, supernatant was harvested by centrifugating the culture at 300 x *g* for 10 min and the pellet discarded. CLEC18A-Fc fusion was purified from the supernatant with an Protein A agarose resin (Gold Biotechnology). Protein A beads in an ethanol solution were centrifuged at 150 x *g* for 5 min to remove the ethanol solution. Pure beads were washed and resuspended with 1X binding buffer (0.02 M Sodium Phosphate, 0.02% sodium azide, pH=7.0) and added into the supernatant together with concentrated binding buffer for a final concentration of 1X binding buffer. Supernatants with beads were incubated overnight 4°C on a shaking plate. After incubation, beads were collected (150 x *g*, 5 min) and washed with binding buffer. Washed beads were transferred to spin columns (G-Biosciences) and centrifuged (100 x *g*, 30 sec) to remove excess buffer. Beads were then mixed with 1 bead volume of elution buffer (100 mM Glycine-HCl, pH=2–3, 0.02% sodium azide) and incubated for 30 sec. The elution was collected into a tube containing neutralization buffer (1 M Tris, pH=9.0, 0.02% sodium azide) by centrifugation at 100 x *g* for 15 sec. The elution procedure was repeated three times, all elution fractions were pooled and the protein concentrations measured with the Pierce™ BCA ProteinAssay Kit (Thermo Fisher Scientific) using the Pierce™ Bovine Gamma Globulin Standard (Thermo Fisher Scientific).

Flow cytometry based cell line interaction screen

All cell lines used in this study were obtained through the ATCC and cultured in DMEM (Thermo Fisher Scientific) supplemented with 10% heat inactivated FCS (Thermo Fisher Scientific), 2 mM L-glutamine (Thermo Fisher Scientific) and 1X Penicillin/Streptomyocin (Thermo Fisher Scientific). All cell lines were routinely checked for mycoplasma and validated with STR profiling.

For binding profiling, cells were harvested at approximately 80% confluency using pre-heated trypsin. Cell culture media was aspirated, cells were washed once with pre-heated PBS and then incubated with trypsin (Thermo Fisher Scientific) until detachment. Cells were harvested with pre-heated DMEM and centrifugated at 500 x *g* for 3 min. Following centrifugation, cells were re-suspended in PBS and incubated with eBioscience™ Fixable Viability Dye eFluor™

780 (Invitrogen) for 20 min at 4°C in the dark. Cells were washed once in TSM buffer (20 mM Tris HCl, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$) and resuspended with 10 µM CLEC18A-Fc fusion or 10 µM Fc control in TSM and incubated for 30 min at 4°C. Following primary lectin incubation, cells were washed and incubated with 1/400 F(ab')2-Goat anti-Mouse IgG (H+L) Secondary Antibody, PE, eBioscience™ for 30 min at 4°C in the dark and then immediately analysed on a BD LSRFortessa™ Cell Analyzer flow cytometer. The data was analyzed in FlowJo (v10.4) and the binding affinity of the CLEC18A-Fc fusion lectin to various cell lines was calculated as the PE signal fold change over the control construct.

CLEC18A-Fc fusion ligand pulldown

Pulldown

PymT and Ac711 cell lines were lysed through repeated freeze/thaw cycles in liquid nitrogen. After lysis, suspensions were centrifuged for 10 min at 2000 x *g* to remove cell debris and the protein concentration of the lysates were determined with the Pierce™ BCA ProteinAssay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions.

For the pulldown, 60 µl of Pierce™ Protein A/G Magnetic Beads (ThermoFisher) suspension was washed three times with PBS on a magnetic rack. After the last wash, beads were re-suspended in 500 µl PBS with 10 µg/µl CLEC18A-Fc fusion and incubated on a shaker for 2h at 4°C. In parallel, PymT and Ac711 lysates were precleared. Again, 60 µl of bead solution was washed three times with PBS and then incubated with 700 µg of lysate in 500 µl PBS. Lysates were incubated on a shaker for 2h at 4°C. After incubation, beads and lysates were washed in PBS and joined together in 500 µl TSM buffer and incubated overnight on a shaker at 4°C. After pulldown, beads were collected and washed 3 times in TSM buffer.

Mass spectrometry analysis of pulldown

Pulldown bead suspension were boiled with 1X Laemmli buffer (Sigma-Aldrich) and run for two minutes into a 12% NuPAGE Bis-Tris gel (Thermo Fisher Scientific) to remove impurities. Samples were subsequently gel extracted and processed further.

Relative peptide amount determination

Final peptide amount was determined by separation of an aliquot of each sample on an LC-UV system equipped with a monolith column. Peptide concentration was calculated based on the peak area of 100 ng of Pierce HeLa protein digestion standard (Thermo Fisher Scientific). Afterwards the peptide solution was frozen at -70°C before further processing.

NanoLC-MS/MS analysis

Samples were run on the nano HPLC system UltiMate 3000 RSLC nano system coupled to the Orbitrap Exploris 480 mass spectrometer, which is equipped with a NanoFlex nanospray source (Thermo Fisher Scientific). Peptides were loaded onto a trap column (Thermo Fisher Scientific, PepMap C18, 5 mm × 300 µm ID, 5 µm particles, 100 Å pore size) at a flow rate of 25 µl/min using 0.1% TFA as mobile phase. After 10 min, the trap column was connected in series with the analytical column (Thermo Fisher Scientific, PepMap C18, 500 mm × 75 µm ID, 2 µm, 100 Å). The analytical column further was connected to PepSep sprayer 1 (Bruker) equipped with a 10 µm ID fused silica electrospray emitter with an integrated liquid junction (Bruker, PN 1893527). Electrospray voltage was set to 2.4 kV. Peptides were eluted using a flow rate of 230 nL/min and a 120 min gradient. The gradient starts with mobile phases A and B: 98% A (water/formic acid, 99.9/0.1, v/v) and 2% B (water/acetonitrile/formic acid, 19.92/80/0.08, v/v/v), increases to 35% B over the next 120 min. This is followed by a gradient increase to 90% B within the next 5 min, stays there for 5 min and decreases again within 2 min back to a gradient of 98% for A and 2% for B to equilibrate at 30°C. The Orbitrap Q-Exactive HF-X mass spectrometer was operated by a mixed MS method which consisted of one full scan (m/z range 380-1 500; 15 000 resolution; AGC target value 3 x 10^6). Maximum injection time was set to 800 ms.

Data analysis

For peptide identification, RAW-files were loaded into Proteome Discoverer (version 2.5.0.400, Thermo Scientific). All created MS/MS spectra were searched using MSAmanda v2.0.0.19924 (Dorfer et al., 2014). As a first step RAW-files were searched against the uniprot_reference_mouse_2023-03-06.fasta (21 928 sequences; 11 722 545 residues), PD_Contaminants_TAGs_v20_tagsremoved.fasta, tags_v11.fasta data bases using following search parameters: The peptide mass tolerance was set to ±10 ppm and the fragment mass tolerance to ±10 ppm. The maximal number of missed cleavages was set to 2. The result was filtered to 1% FDR on protein level using the Percolator algorithm integrated in Thermo Proteome Discoverer. Additionally a sub-database was generated for further processing. As a second step RAW-files were searched against the created sub-database. The following search parameters were used: Iodoacetamide derivative on cysteine was set as a fixed modification, oxidation on methionine, phosphorylation on serine, threonine and tyrosine, deamidation on asparagine and glutamine, pyro-glu from q on peptide N-terminal glutamine, acetylation on protein N-Terminus were set as variable modifications. Monoisotopic masses were searched within unrestricted protein masses for tryptic enzymatic specificity. The peptide mass tolerance

again was set to ±10 ppm and the fragment mass tolerance to ±10 ppm. Maximal number of missed cleavages was set to 2. The result was filtered to 1% FDR on protein level using Percolator algorithm integrated in Thermo Proteome Discoverer. Additional high quality filtering by setting a minimum MS Amanda Score of 150 on PSM level was applied.

Protein areas were quantified using IMP-apQuant [57] by summing unique and razor peptides and applying intensity-based absolute quantification (iBAQ) [58] with subsequent normalisation based on the MaxLFQ algorithm [59]. Proteins were filtered to be identified by a minimum of 2 PSMs in at least one sample and identified proteins were pre-filtered to contain at least three quantified peptide groups. Pulldown hits are shown in **Supplementary Table 1**.

AlphaFold2 Multimer interaction screen

AlphaFold2 Multimer [20-22] was impleneted and used to predict interactions between CLEC18A, CLEC18B and CLEC18C and putative interaction partners. A list of 379 kidney and clear cell renal cell carcinoma specific proteins was generated by: extracting (i) proteins that in the literature are defined as proximal tubule specific (expression site of CLEC18A) [18, 60, 61], (ii) proteins that in the human protein atlas [62] are defined as kidney enriched and/or only found in kidney, and (iii) proteins that were angiotensin related [63], were pathway signature associated [64], had high mutational frequency [65], were associated with a ccRCC inflammatory state [66], had high recurrent repeat expansions in ccRCC [67], or proteins that in the cBioPortal [68] were mutated in more than 5% of all ccRCC cases and/or correlated (Spearman's correlation > 0.5) with *CLEC18A* expression. After assembling the list of targets redundancies were removed. CLEC18A, CLEC18B and CLEC18C were all screened against the list of proteins resulting in 1110 pairwise PPI predictions using a custom script on a local CPU and GPU cluster using MMseqs (https://github.com/soedinglab/MMseqs2) for local Multiple Sequence Alignment creation and colabfold (https://github.com/sokrypton/ColabFold) for structure prediction with 5 models per prediction and omitting structure relaxation.

Hits from the CLEC18A-Fc fusion pulldown experiment were analyzed in the same way.

GlycoArray

Glycan arrays were printed as previously described by [69]. In brief, 254 glycans (**Supplementary Table 2**) were printed using an Arrayjet Mercury Microarray printer at a concentration of 1000 µM onto sciCHIP epoxy slides (Scienion). Slides were neutralised, rinsed in 100% ethanol and stored at 4ºC under vacuum until use. For all array work, proteins

are diluted in array PBS (PBS, 1.8 mM $MgCl_2$, 1.8 mM $CaCl_2$, pH 7.4), filtered through 0.22 µm filter and degassed prior to use. 2 µg of Fc control and CLEC18A-Fc fusion were incubated with AlexaFluor647 rabbit anti-mouse and goat anti-rabbit IgG (Invitrogen) at a molar ratio of 4:2:1 in array PBS and preincubated on ice for 10 mins at a total volume of 65 µl. A 65 µl geneframe was placed onto the slide surface and the incubation mix was subsequently added and sealed with a coverslip. The sample was incubated for 15 mins at room temperature in the dark. After the incubation time, the slide was immersed in array PBS and the geneframe and coverslip were removed. The slides were washed briefly for two mins with gentle shaking. The slides were dried by centrifugation at 200 x *g* for 4 min. The slide was then scanned using the 647 nm laser and 677/45 filter, with low laser power and 50% PMT gain settings. The acquired image was then analysed using the Mapix software, overlaying the image with the map (Gal) file. The experiment was repeated three times, and binding was classified as positive when the average RFU (relative fluorescence units) of a specific structure had a value above mean background (defined as the average background fluorescence plus 3 standard deviations) and had a P value of < 0.005 (student's t-test). A pre-experimental image was acquired to rule out any autofluorescence of the printed glycans.

TCGA kidney cancer data analysis

TCGA RNA sequencing data from tumor and normal tissue for TCGA-KIRC, TCGA-KIRP and TCGA-KICH was downloaded through the GDC portal. Differential expression analysis between tumor and matched normal tissue was performed with DESeq2 (RRID:SCR_015687, v1.22.2). Significantly upregulated or downregulated genes were defined as genes with ≥10 counts, and Benjamini-Hochberg adjusted Wald p values of $p < 0.05$.

***Clec18a* knockout generation**

gRNAs for murine *Clec18a* were designed using CHOPCHOP [70], synthesized with appropriate overhangs and cloned into the pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid [71]. Cells were transfected using the MaxCyte® STX electroporation system. For this, 10^6 cells per cell line were prepared by washing with PBS and HyClone Electroporation Buffer (MaxCyte® once. Then, RAG cells were resuspended with 100 µl HyClone Electroporation Buffer (MaxCyte) and 20 µL of PX459 plasmid containing the *Clec18a* gRNA. The mix containing cells and plasmid was transferred into an OC-100 cuvette and electroporated using the "Renca" program from MaxCyte. Electroporated cells were then transferred into a 10 cm dish and

incubated for 30 min at 37°C. Then, they were resuspended with 10 ml of supplemented DMEM media and incubated over night at 37°C. The next day for selecting transfected cells 2 µg/ml of puromycin was added for 24 hours. After seeded cells reached confluency single cell clones were seeded in two 96-well plates per cell line using the FACS Aria III cell sorter. Individual clones were a region flanking the mutation site amplified using PCR and the PCR fragment was purified and subjected to Sanger sequencing (Vienna BioCenter inhouse service). The editing efficiency was assessed using TIDE analysis [72]. Clones with a predicted efficiency of >95% and with a frameshift mutation were expanded and validated using RT-qPCR.

Clec18a overexpression generation

E0771 cells overexpressing *Clec18a* were generated using a lentiparticle construct pre-assembled by OriGene. E0771 cells were transduced with lentiparticles containing containing *Clec18a* overexpression plasmid or an empty control plasmid, both with puromycin resistance. According to the manufacturer's instructions, on the first day 0.24 x 10^6 cells were seeded in each well of a 6-well plate, whereas two wells were seeded per cell line, one for each condition and incubated for 20 hours at 37°C. The next day for viral infection of the cells lentiviral particles were added according to a multiplicity of infection (MOI) of 10, cell culture media and 8 µg/mL of polybrene. Cells were then incubated for 18 hours at 37°C. Afterwards lentivirus containing medium was removed and exchanged with normal complete DMEM. After cell line generation, E0771 cells were maintained in media containing 1 µg/ml puromycin.

Mouse studies

Mus musculus

All mice used in this study were bred and maintained at the Comparative Medicine Mousehouse at the Vienna BioCenter. The mice were maintained in pathogen free environments with regular health screenings in a 12 hour light-dark cycle. Food and water was provided *ab libitum*. All mouse experiments were approved by the Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFW, project 2023-0.517.898), and carried out according to EU-directive 2010/63/EU.

Subcutaneous tumor injection

Clear cell renal cell carcinoma cells (RAG) were injected subcutaneously into the left flank and breact cancer cells (E0771) were injected orthotopically into the right inguinoabdominal mammary fat pad under anaesthesia with xylasol and ketasol. Per mouse, 500000 RAG cells or 250000 E0771 cells were injected. RAG cells did not grow properly in wildtype syngenic BALB/c mice so all mouse experiments were performed in *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ immunodeficient mice (The Jackson Laboratory). Tumor size was measured every 3-4 days and the volume was calculated as $V = \frac{R \times r^2}{2}$ where R is the longest diameter and r is short diamater.

Immunohistochemistry

Tumors were harvested as mice reached human endpoint and fixed in neutral buffered formalin (Sigma-Aldrich) for 24-48h, room temperature. Following fixation, tumors were washed in PBS and embedded in paraffin and sectioned at 2 µm thickness. Sections were mounted on glass slides (Permaflex plus adhesive slides, Leica). Slides were deparaffinized and rehydrated with a Gemini AS automatic stainer (Fisher Scientific). Rehydrated slides were subjected to heat-induced epitope retrieval with pH 6 sodium citrate buffer, followed by endogenous peroxidase inactivation with 3% H_2O_2 (Sigma-Aldrich) and subsequently blocked in 5% BSA in TBS-T with TBS washes inbetween each step. Slides were incubated with primary rabbit anti-Ly6G or rabbit anti-F4/80 antibodies (Cell Signaling Technology) for 2h at room temperature and subsequently treated with the a rabbit detection system (DCS) according to the manufacturers instructions. Chromogenic reaction was induced with a DAB substrate kit (Abcam) and hematoxylin counter taining was performed with the Gemini AS automatic stainer. Slides were mounted with cover slips by Tissue-TEK GLC (Sakura Finetek) and scanned with Slide Scanner Pannoramic 250 (3DHistech Ltd).

RNAScope *in situ* hybridization

Brains were extracted from 10 week old wildtype C57B6J mice and fixed overnight in 4% paraformaldehyde followed by dehydration and paraffin embedding. Following embedding, 3.5 µm sections were cut throughout the brain and *in situ* hybridized with a custom designed RNAScope probe for murine *Clec18a* by ACD biosciences. *In situ* hybridization was performed according to manufacturers instructions. The sections were pretreated for 15 minutes with Target Retrieval Solution and for 30 minutes with Protease Plus. After hybridization, the mRNA signal was amplified and visualized with an alkaline phosphatase–based red substrate.

Statistical analysis

All statistical analysis were performed in Prism 10 (GraphPad) or in RStudio (Posit). When comparing large omics datasets p values were calculated with limma-moderated Benjamini–Hochberg-corrected two-sided t-test after data processing for proteomics, and Benjamini-Hochberg adjusted Wald p values for TCGA data. For comparisons of individual markers between groups, the distribution of the data was initially determined by Shapiro-Wilk normality test. Statistical testing was then performed with Mann-Whitney U-tests (non-normal data) or two-sided t-tests (normal data). Tumor time course significance was assessed with two-way ANOVA with Šídák's multiple comparisons test.

Disclosure statement

The authors have no conflicts of interest to declare.

Author contribution

G.J. and J.M.P conceived the study. G.J. performed and designed experiments with input and help from all co-authors as follows: M.H. with cell line generation, mouse experiments and AlphaFold2 Multimer screens, L.H-T with glycoarray experiments, S.M. and T.O. with pulldowns and mass spectrometry, I.S. with mouse experiments, D.H. with recombinant CLEC18A-Fc generation, M.N. with TCGA analysis and A.S. with phylogeny analysis.

Acknowledgements

We thank all members of our laboratories for their support and help. The RNA-FISH and immunohistochemistry was performed by the Histology Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria. Proteomics analyses were performed by the Proteomics Facility at IMP/IMBA/GMI using the VBCF instrument pool. The YUMM cell lines were a gift from the Obenauf lab. G.J. is supported by a DOC fellowship from the Austrian Academy of Sciences. S.M. received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 841319 and the ESPRIT-Programme of the Austrian Science Fund (FWF, Project number: ESP 166). The lab of J.M.P. received funding from the Medical University of Vienna, the Austrian Academy of Sciences, the T. von Zastrow foundation, the Fundacio La Marato de TV3 (202125-31), and the Canada 150 Research Chairs Program F18-01336. We also gratefully acknowledge funding by the German Federal Ministry of Education and

Research (BMBF) under the project "Microbial Stargazing - Erforschung von Resilienzmechanismen von Mikroben und Menschen" (Ref. 01KX2324).

Data availability

Data for the glycoarray and hits from the pulldown experiment are available as supplementary tables. All other raw data is available from the lead authors upon request.

Conflict of interest

The authors have no conflict of interest to declare

References

1. Brown, G.D., J.A. Willment, and L. Whitehead, *C-type lectins in immunity and homeostasis.* Nature Reviews Immunology, 2018. **18**(6): p. 374-389.
2. Geijtenbeek, T.B. and S.I. Gringhuis, *C-type lectin receptors in the control of T helper cell differentiation.* Nature Reviews Immunology, 2016. **16**(7): p. 433-448.
3. Mayer, S., M.-K. Raulf, and B. Lepenies, *C-type lectins: their network and roles in pathogen recognition and immunity.* Histochemistry and cell biology, 2017. **147**: p. 223-237.
4. Drickamer, K. and M.E. Taylor, *Recent insights into structures and functions of C-type lectins in the immune system.* Current opinion in structural biology, 2015. **34**: p. 26-34.
5. Powlesland, A.S., et al., *Widely divergent biochemical properties of the complete set of mouse DC-SIGN-related proteins.* Journal of biological chemistry, 2006. **281**(29): p. 20440-20449.
6. Akiyama, J., et al., *Characterization of the mouse collectin gene locus.* American journal of respiratory cell and molecular biology, 1999. **21**(2): p. 193-199.
7. Seifert, L., et al., *The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression.* Nature, 2016. **532**(7598): p. 245-249.
8. Daley, D., et al., *Dectin 1 activation on macrophages by galectin 9 promotes pancreatic carcinoma and peritumoral immune tolerance.* Nature medicine, 2017. **23**(5): p. 556-567.
9. Ding, D., et al., *C-type lectins facilitate tumor metastasis.* Oncology letters, 2017. **13**(1): p. 13-21.
10. Lanier, L.L., *NKG2D Receptor and Its Ligands in Host Defense.* Cancer Immunology Research, 2015. **3**(6): p. 575-582.
11. Kimura, Y., et al., *The innate immune receptor Dectin-2 mediates the phagocytosis of cancer cells by Kupffer cells for the suppression of liver metastasis.* Proceedings of the National Academy of Sciences, 2016. **113**(49): p. 14097-14102.
12. Seifert, L., et al., *Dectin-1 regulates hepatic fibrosis and hepatocarcinogenesis by suppressing TLR4 signaling pathways.* Cell reports, 2015. **13**(9): p. 1909-1921.

13. Huang, Y.-L., et al., *Human CLEC18 gene cluster contains C-type lectins with differential glycan-binding specificity.* Journal of Biological Chemistry, 2015. **290**(35): p. 21252-21263.
14. Tsai, T.-Y., et al., *The human C-type lectin 18 is a potential biomarker in patients with chronic hepatitis B virus infection.* Journal of biomedical science, 2018. **25**: p. 1-10.
15. Liao, T.-L., et al., *Association of C-type lectin 18 levels with extrahepatic manifestations in chronic HCV infection.* Scientific Reports, 2018. **8**(1): p. 17287.
16. Huang, Y.-L., et al., *Endosomal TLR3 co-receptor CLEC18A enhances host immune response to viral infection.* Communications Biology, 2021. **4**(1): p. 229.
17. Cheng, L., et al., *Transgenic expression of human C-type lectin protein CLEC18A reduces dengue virus type 2 infectivity in Aedes aegypti.* Frontiers in Immunology, 2021. **12**: p. 640367.
18. Stewart, B.J., et al., *Spatiotemporal immune zonation of the human kidney.* Science, 2019. **365**(6460): p. 1461-1466.
19. Varki, A., et al., *Essentials of Glycobiology [internet].* 2015.
20. Evans, R., et al., *Protein complex prediction with AlphaFold-Multimer.* biorxiv, 2021: p. 2021.10. 04.463034.
21. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
22. Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nature methods, 2022. **19**(6): p. 679-682.
23. Sandoval, D.R., et al., *Proteomics-based screening of the endothelial heparan sulfate interactome reveals that C-type lectin 14a (CLEC14A) is a heparin-binding protein.* Journal of Biological Chemistry, 2020. **295**(9): p. 2804-2821.
24. Li, Q., *The multiple roles of C-type lectin receptors in cancer.* Frontiers in Oncology, 2023. **13**.
25. Dambuza, I.M. and G.D. Brown, *C-type lectins in immunity: recent developments.* Current opinion in immunology, 2015. **32**: p. 21-27.
26. Sancho, D., et al., *Identification of a dendritic cell receptor that couples sensing of necrosis to immunity.* Nature, 2009. **458**(7240): p. 899-903.
27. Schreibelt, G., et al., *The C-type lectin receptor CLEC9A mediates antigen uptake and (cross-) presentation by human blood BDCA3+ myeloid dendritic cells.* Blood, The Journal of the American Society of Hematology, 2012. **119**(10): p. 2284-2292.
28. Yan, H., et al., *Targeting C-type lectin receptors for cancer immunity.* Frontiers in immunology, 2015. **6**: p. 146351.
29. Garcia-Vallejo, J.J. and Y. van Kooyk, *The physiological role of DC-SIGN: a tale of mice and men.* Trends in immunology, 2013. **34**(10): p. 482-486.
30. Chang, C.-M., W.-C. Chang, and S.L. Hsieh, *Characterization of the genetic variation and evolutionary divergence of the CLEC18 family.* Journal of Biomedical Science, 2024. **31**(1): p. 53.
31. Bülow, R.D. and P. Boor, *Extracellular matrix in kidney fibrosis: more than just a scaffold.* Journal of Histochemistry & Cytochemistry, 2019. **67**(9): p. 643-661.

32. Rabelink, T.J., et al., *Heparanase: roles in cell survival, extracellular matrix remodelling and the development of kidney disease.* Nature Reviews Nephrology, 2017. **13**(4): p. 201-212.
33. Aikawa, T., et al., *Glypican-1 modulates the angiogenic and metastatic potential of human and mouse cancer cells.* The Journal of clinical investigation, 2008. **118**(1): p. 89-99.
34. Yao, W., et al., *Syndecan 1 is a critical mediator of macropinocytosis in pancreatic cancer.* Nature, 2019. **568**(7752): p. 410-414.
35. Schulz, G.B., et al. *Prognostic value and association with epithelial-mesenchymal transition and molecular subtypes of the proteoglycan biglycan in advanced bladder cancer*. in *Urologic Oncology: Seminars and Original Investigations*. 2019. Elsevier.
36. Zhang, Y., et al., *Enhanced PAPSS2/VCAN sulfation axis is essential for Snail-mediated breast cancer cell migration and metastasis.* Cell Death & Differentiation, 2019. **26**(3): p. 565-579.
37. Dudani, S., et al., *Evaluation of clear cell, papillary, and chromophobe renal cell carcinoma metastasis sites and association with survival.* JAMA network open, 2021. **4**(1): p. e2021869-e2021869.
38. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic acids research, 1997. **25**(17): p. 3389-3402.
39. *UniProt: the universal protein knowledgebase in 2021.* Nucleic acids research, 2021. **49**(D1): p. D480-D489.
40. *Database resources of the national center for biotechnology information.* Nucleic acids research, 2018. **46**(D1): p. D8-D13.
41. Katoh, K. and H. Toh, *Recent developments in the MAFFT multiple sequence alignment program.* Briefings in bioinformatics, 2008. **9**(4): p. 286-298.
42. Minh, B.Q., et al., *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.* Mol Biol Evol, 2020. **37**(5): p. 1530-1534.
43. Kalyaanamoorthy, S., et al., *ModelFinder: fast model selection for accurate phylogenetic estimates.* Nature methods, 2017. **14**(6): p. 587-589.
44. Hoang, D.T., et al., *UFBoot2: improving the ultrafast bootstrap approximation.* Molecular biology and evolution, 2018. **35**(2): p. 518-522.
45. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.* Systematic biology, 2010. **59**(3): p. 307-321.
46. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.* Nucleic acids research, 2021. **49**(W1): p. W293-W296.
47. Livingstone, C.D. and G.J. Barton, *Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation.* Bioinformatics, 1993. **9**(6): p. 745-756.
48. Erdős, G., M. Pajkos, and Z. Dosztányi, *IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation.* Nucleic acids research, 2021. **49**(W1): p. W297-W303.

49. Gouy, M., S. Guindon, and O. Gascuel, *SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building.* Molecular biology and evolution, 2010. **27**(2): p. 221-224.
50. Varadi, M., et al., *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models.* Nucleic acids research, 2022. **50**(D1): p. D439-D444.
51. Pettersen, E.F., et al., *UCSF ChimeraX: Structure visualization for researchers, educators, and developers.* Protein science, 2021. **30**(1): p. 70-82.
52. Golicz, A., et al., *AACon: a fast amino acid conservation calculation service.* Submitted paper, 2018.
53. Biology, C.S.-C., et al., *CZ CELLxGENE Discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data.* bioRxiv, 2023: p. 2023.10. 30.563174.
54. Megill, C., et al., *Cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices.* bioRxiv, 2021: p. 2021.04. 05.438318.
55. Hao, Y., et al., *Dictionary learning for integrative, multimodal and scalable single-cell analysis.* Nature biotechnology, 2024. **42**(2): p. 293-304.
56. Hoffmann, D., et al., *Identification of lectin receptors for conserved SARS-CoV-2 glycosylation sites.* The EMBO Journal, 2021. **40**(19): p. e108375.
57. Doblmann, J., et al., *apQuant: accurate label-free quantification by quality filtering.* Journal of Proteome Research, 2018. **18**(1): p. 535-541.
58. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-342.
59. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Molecular & cellular proteomics, 2014. **13**(9): p. 2513-2526.
60. Gerhardt, L.M., et al., *Single-nuclear transcriptomics reveals diversity of proximal tubule cell states in a dynamic response to acute kidney injury.* Proceedings of the National Academy of Sciences, 2021. **118**(27): p. e2026684118.
61. Park, J., et al., *Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease.* Science, 2018. **360**(6390): p. 758-763.
62. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
63. Khanna, P., et al., *ACE2 abrogates tumor resistance to VEGFR inhibitors suggesting angiotensin-(1-7) as a therapy for clear cell renal cell carcinoma.* Science translational medicine, 2021. **13**(577): p. eabc0170.
64. Tun, H.W., et al., *Pathway signature and cellular differentiation in clear cell renal cell carcinoma.* PloS one, 2010. **5**(5): p. e10696.
65. Sato, Y., et al., *Integrated molecular analysis of clear-cell renal cell carcinoma.* Nature genetics, 2013. **45**(8): p. 860-867.
66. Rappold, P., L. Vuong, and J. Leibold, *A targetable myeloid inflammatory state governs disease recurrence in clear-cell renal cell carcinoma. Cancer Discov. 2022; 12 (10): 2308-2329*.

67. Erwin, G.S., et al., *Recurrent repeat expansions in human cancer genomes.* Nature, 2023. **613**(7942): p. 96-102.
68. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Science signaling, 2013. **6**(269): p. pl1-pl1.
69. Day, C.J., et al., *Differential carbohydrate recognition by Campylobacter jejuni strain 11168: influences of temperature and growth conditions.* PloS one, 2009. **4**(3): p. e4927.
70. Labun, K., et al., *CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing.* Nucleic acids research, 2019. **47**(W1): p. W171-W174.
71. Ran, F., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature protocols, 2013. **8**(11): p. 2281-2308.
72. Brinkman, E.K., et al., *Easy quantitative assessment of genome editing by sequence trace decomposition.* Nucleic acids research, 2014. **42**(22): p. e168-e168.

Figures and Figure legends

Figure 1.

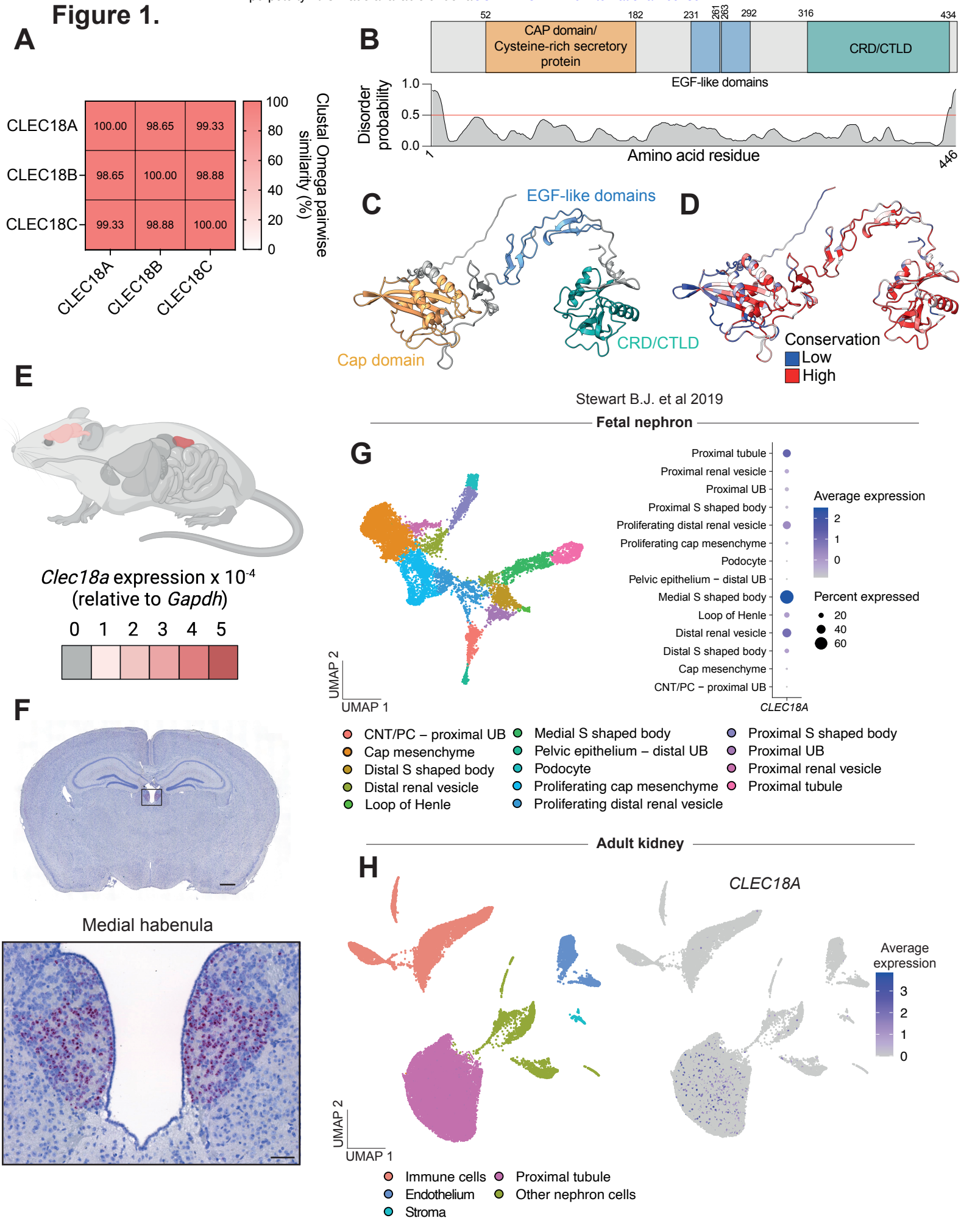


Figure 1. The CLEC18 family are highly conserved proteins and gene expression is restricted to the brain and kidneys

A. Similarity matrix between the amio acid sequences of human CLEC18A, CLEC18B and CLEC18C calculated through Clustal Omega pairwise similarity.

B. Human CLEC18A protein domains and disorder. Disorder is calculated using PrDOS.

C. AlphaFold2 prediction of human CLEC18A structure with protein domains highlighted.

D. AlphaFold2 prediction of human CLEC18A structure with overlayed conservation scored calculated by the predicted local distance difference test (pLDDT).

E. *Clec18a* expression in all organs of mice measured with RT-qPCR. Expression data is shown as an organ heatmap overlayed on a mouse. Expression was predominantely detected in the kidney and brain. Schematic created using BioRender.com.

F. RNAscope *in situ* hybridization of *Clec18a* on coronal sections of mouse brains. Positive cells display a red color in the insert of the medial habenula. Scale bar for the full coronal brain section = 1 mm. Scale bar for the insert = 100 μm.

G. Re-analysis of scRNA-seq data from the kidney cell atlas [18] showing clustering of fetal nephron cells and detection of *CLEC18A* in the medial S shaped body and proximal tubules of fetal nephrons as a bubble plot. Abbreviations: UMAP = Unifold Manifold Approximation and Projection, UB = Ureteric bud, CNT = Connecting tubule, PC = Principal cell.

H. Re-analysis of scRNA-seq data from the kidney cell atlas [18] showing detection of *CLEC18A* in the proximal tubule of adult kidneys. Abbreviations: UMAP = Unifold Manifold Approximation and Projection.

Figure 2.

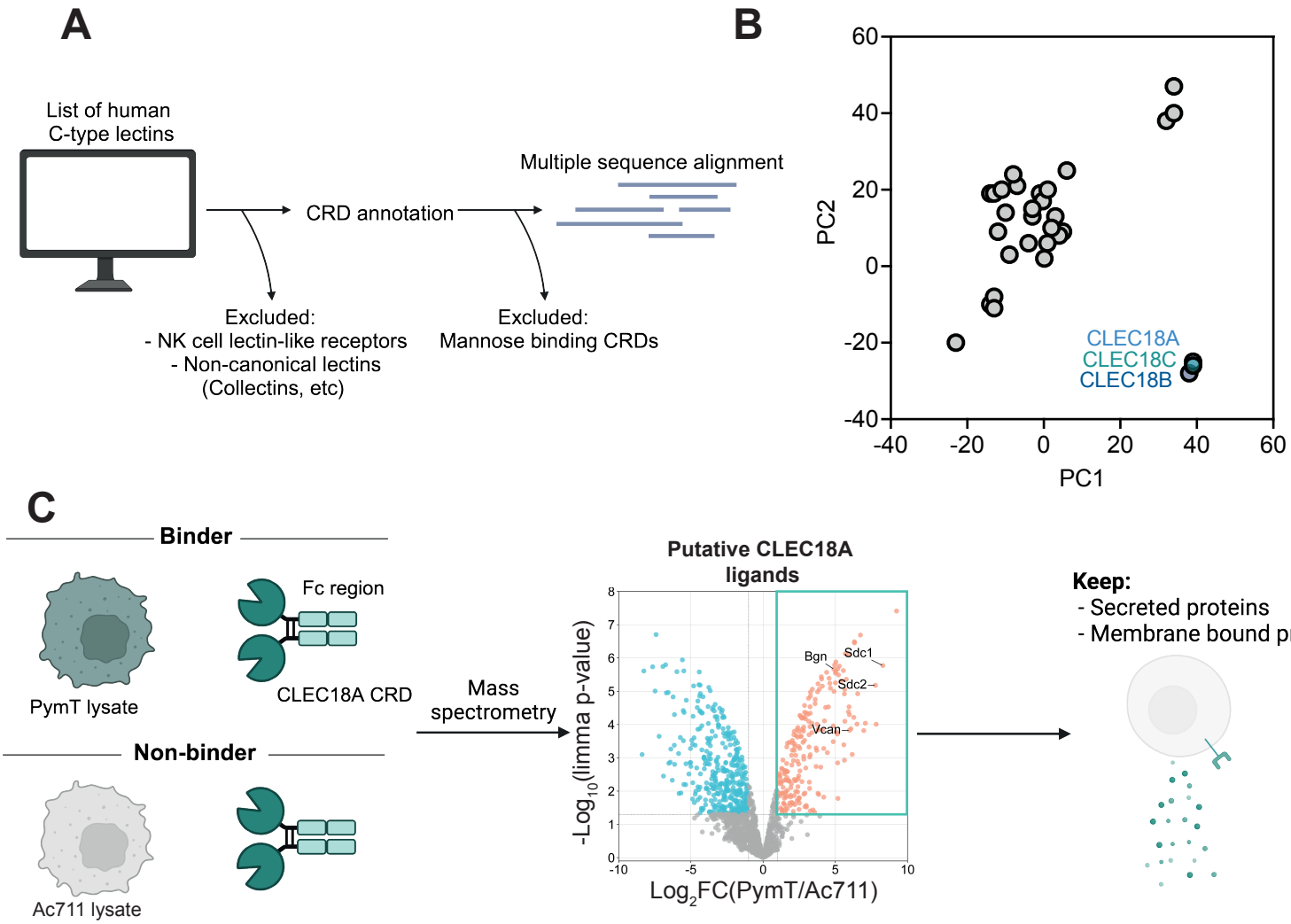


Figure 2. CLEC18A interacts with proteoglycans

A. Schematic of CRD annotation of all human C-type lectins and multiple sequence alignment of those sequences. Abbreviations: CRD = Carbohydrate recognition domain. Schematic created using BioRender.com.

B. Principal component analysis using JalView on the generated multiple sequence alignment of the human CRD sequences. Abbreviations: PC = Principal component

C. Schematic of CLEC18A-Fc fusion pulldown strategy against two cell lines, one with high binding affinity for CLEC18A (PymT) and one with no binding affinity to CLEC18A (Ac711). Mass spectrometry was performed on pulldowns and putative CLEC18A ligands were filtered for proteins which are either secreted or bound to the membrane. Volcano plot shows upregulated (red) and downregulated (blue) proteins in PymT lysates (binders) vs Ac711 lysates (non-binders). Some proteoglycans are highlighted as interaction partners. Abbreviations: CRD = Carbohydrate recognition domain, Bgn = Biglycan, Sdc1 = Syndecan 1, Sdc2 = Syndecan 2, Vcan = Vesican. P values were calculated with limma-moderated Benjamini–Hochberg-corrected two-sided t-test. Abbreviations: FC = Fold change. Schematic created using BioRender.com.

D. Filtered secreted and membrane bound ligands hits from **C** shown as a waterfall plot. Proteoglycan and proteoglycan related proteins are highlighted. Abbreviations: FC = Fold change.

Figure 3.

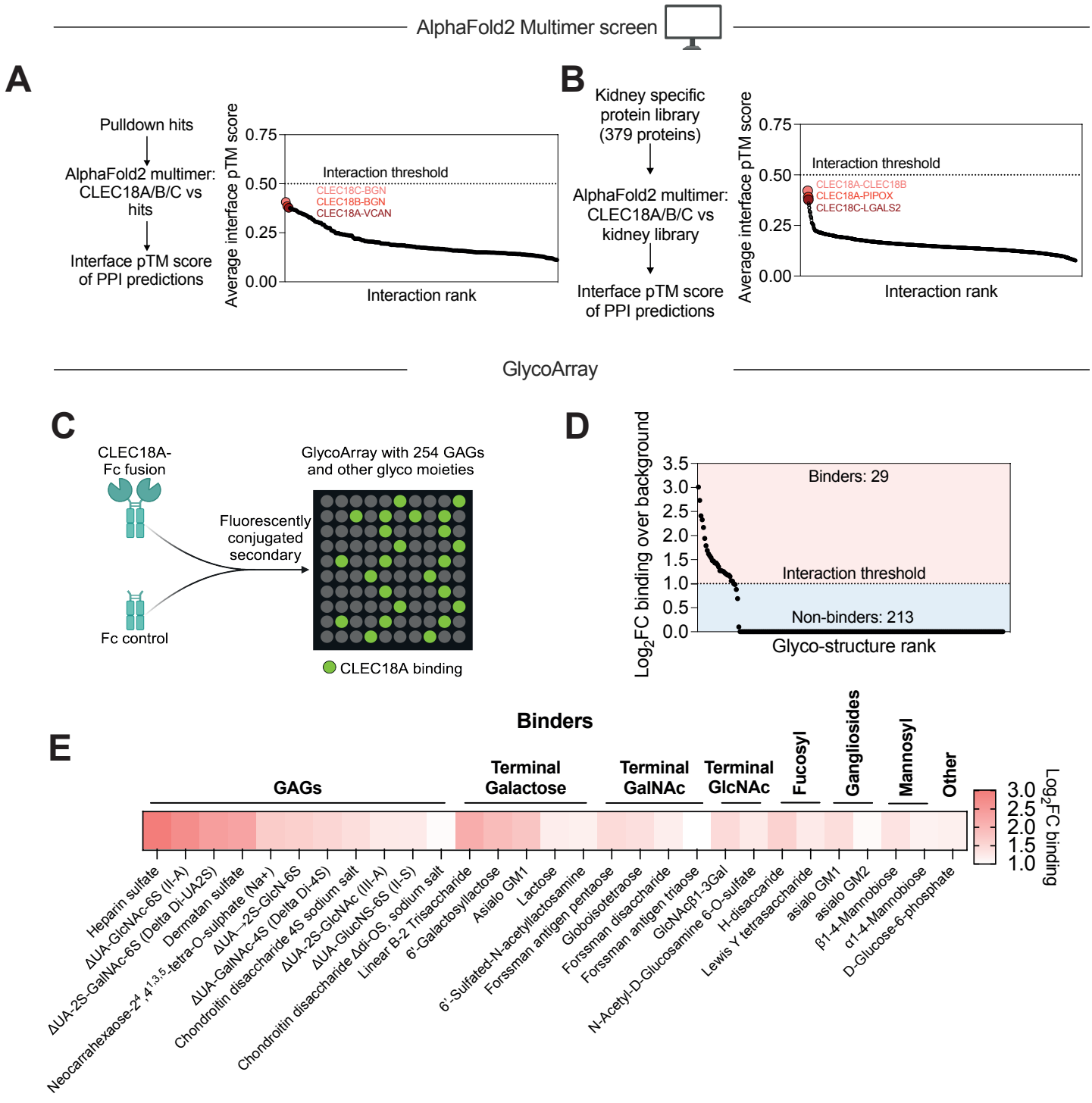


Figure 3. CLEC18A interacts with glycosaminoglycans, and not amino acid residues, on proteoglycans

A. AlphaFold2 Multimer screen of CLEC18A, CLEC18B and CLEC18C against the top ligand candidates in the CLEC18A-Fc fusion pulldown. The interaction threshold of interface pTM > 0.5 is not reached for any of the top hits indicating that these are not protein-protein interactions. The three highest scoring predictions are highlighted. Abbreviations: pTM = predicted template modelling, PPI = Protein-protein interaction.

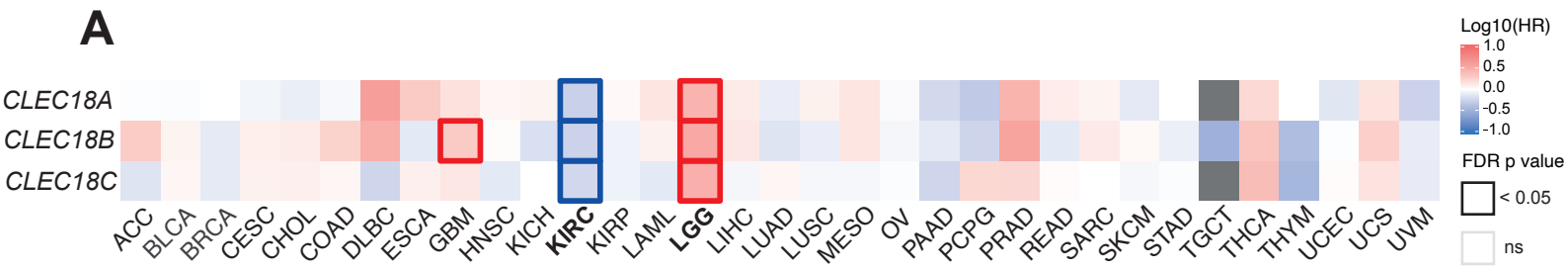
B. AlphaFold2 Multimer screen of CLEC18A, CLEC18B and CLEC18C against a library of 379 kidney and clear cell renal cell carcinoma specific proteins. The three highest scoring predictions are highlighted. Abbreviations: pTM = predicted template modelling, PPI = Protein-protein interaction.

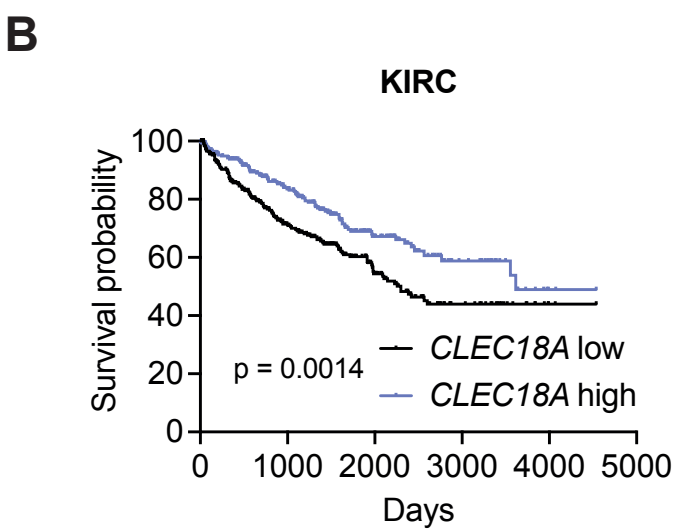
C. Schematic for detection of glyco-ligands for CLEC18A using a GlycoArray with 260 GAGs and other glycol moieties. Full list of glyco structures included in the array are available in Supplementary Table 1. Abbreviations: GAG = Glycosaminoglycan.

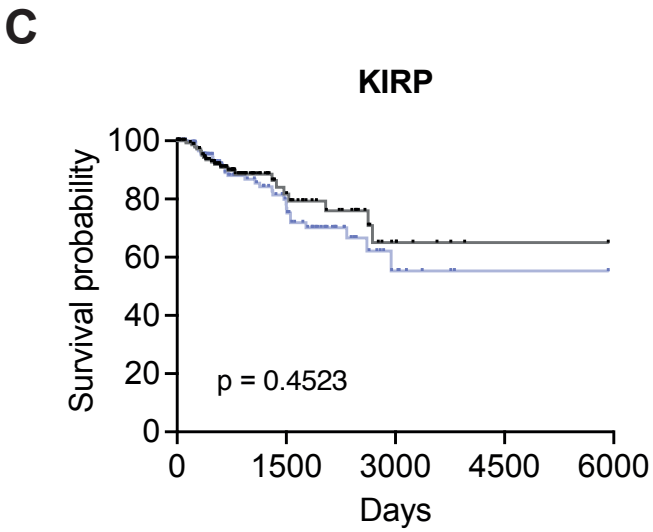
D. Waterfall plot of potential glyco-ligands for CLEC18A from the GlycoArray. Structures which displayed binding to the Fc control have been removed. Abbreviations: FC = Fold change.

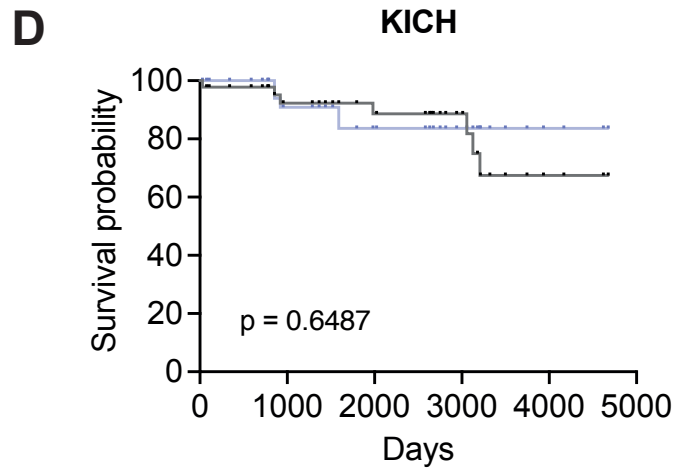
E. Heatmap of the 29 binders which were shown to interact with CLEC18A. Abbreviations: GAG = Glycosaminoglycan, GalNAc = N-Acetylgalactosamine, GlcNAc = N-Acetylglucosamine, FC = Fold change.

Figure 4.









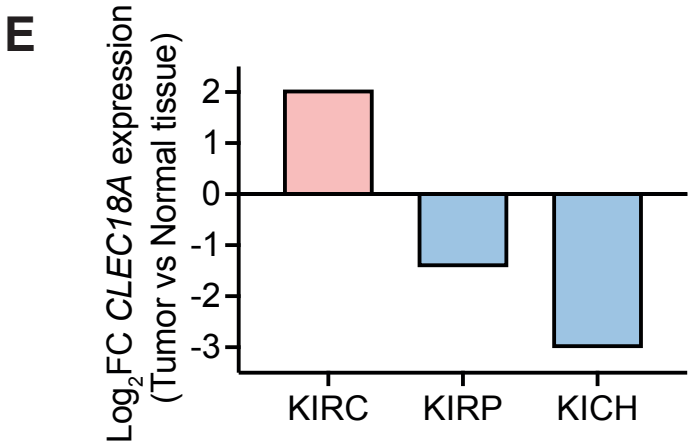


Figure 4. CLEC18A expression correlates with survival in clear cell renal cell carcinoma

A. Survival heatmap for high vs low expression (high and low expression attributed based on the median) of *CLEC18A, CLEC18B* and *CLEC18C* for all cancer types in the The Cancer Genome Atlas shown as the hazard ratio. A think outline indicates significant hazard ratios. Cancer types where all three CLEC18 paralogs contribute to a significant shift in survival based on high and low expression are shown in bold; Clear cell renal cell carcinoma (KIRC) and Low grade glioma (LGG). Abbreviations: HR = Hazard ratio, FDR = False discovery rate.

B. Kaplan meier survival curve for high vs low *CLEC18A* expression (high and low expression attributed based on the median) in KIRC. P value calculated with a Log rank test. Abbreviations: KIRC = Clear cell renal cell carcinoma.

C. Kaplan meier survival curve for high vs low *CLEC18A* expression (high and low expression attributed based on the median) in KIRP. P value calculated with a Log rank test. Abbreviations: KIRP = Papillary renal cell carcinoma.

D. Kaplan meier survival curve for high vs low *CLEC18A* expression (high and low expression attributed based on the median) in KICH. P value calculated with a Log rank test. Abbreviations: KICH = Chromophobe renal cell carcinoma.

E. Expression comparison of *CLEC18A* between tumor and normal tissue for the three kidney cancer types that are present in The Cancer Genome Atlas. Abbreviations: KIRC = Clear cell renal cell carcinoma, KIRP = Papillary renal cell carcinoma, KICH = Chromophobe renal cell carcinoma, FC = Fold change.

Figure 5.

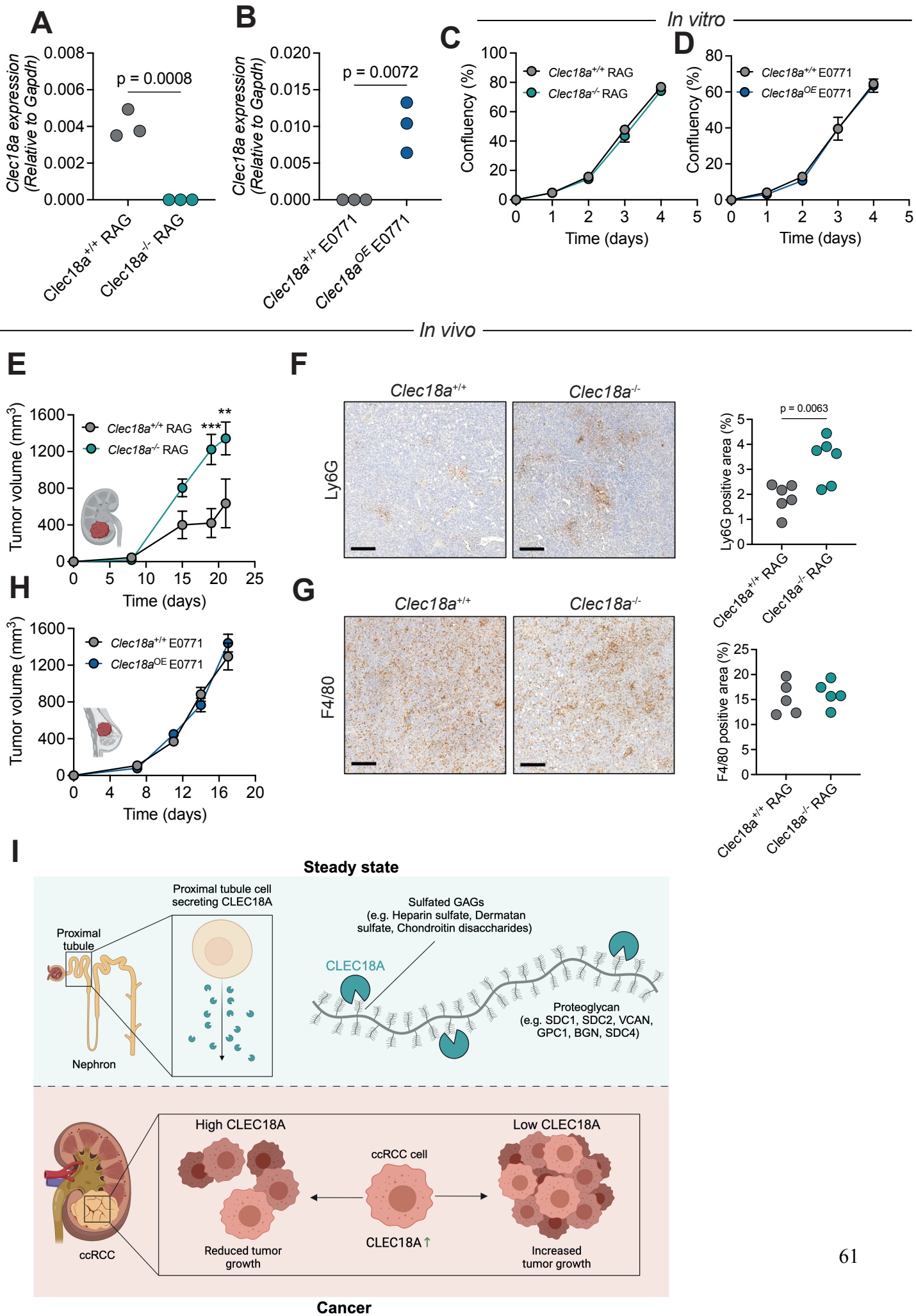


Figure 5. CLEC18A expression restricts tumor progression in murine models of ccRCC

A. Validation of *Clec18a* knockout in *Clec18a*$^{-/-}$ RAG cells. P value calculated with a two-sided Student's t-test.

B. Validation of *Clec18a* overexpression in *Clec18a*^{OE} E0771 cells. P value calculated with a two-sided Student's t-test. Abbreviations: OE = overexpression.

C. *In vitro* growth curves of *Clec18a*$^{+/+}$ and *Clec18a*$^{-/-}$ RAG cells. P values calculated assessed with two-way ANOVA with Šídák's multiple comparisons test.

D. *In vitro* growth curves of *Clec18a*$^{+/+}$ (empty vector control) and *Clec18a*^{OE} E0771 cells. P values calculated assessed with two-way ANOVA with Šídák's multiple comparisons test.

E. Tumor growth curve kinetics of *Clec18a*$^{+/+}$ and *Clec18a*$^{-/-}$ clear cell renal cell carcinoma cell lines (RAG) in *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice. P values calculated assessed with two-way ANOVA with Šídák's multiple comparisons test.

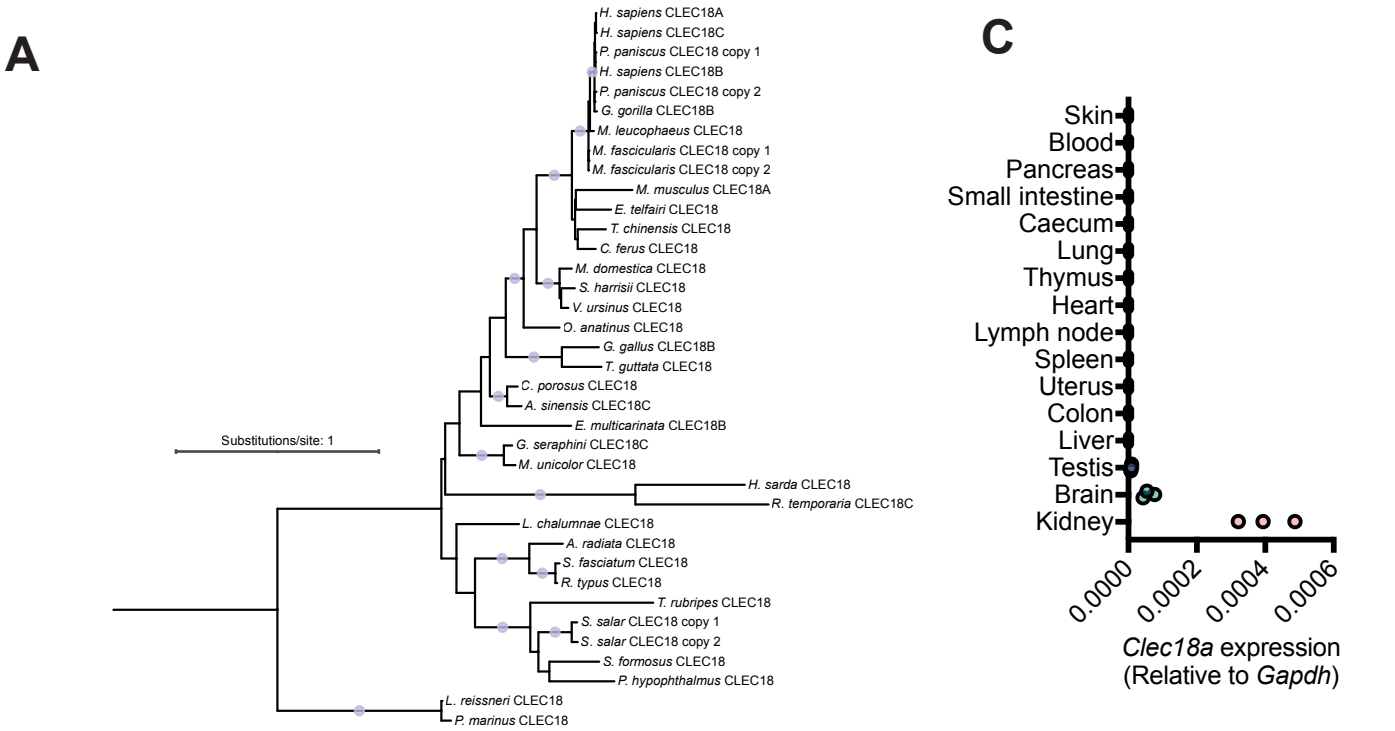
F. Assessment of neutrophil infiltration through Ly6G staining in *Clec18a*$^{+/+}$ and *Clec18a*$^{-/-}$ tumors subcutaneously injected into *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice. Scale bars = 100 μm. P value calculated with a two-sided Student's t-test.

G. Assessment of macrophage infiltration through F4/80 staining in *Clec18a*$^{+/+}$ and *Clec18a*$^{-/-}$ tumors subcutaneously injected into *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice. Scale bars = 100 μm. P value calculated with a two-sided Student's t-test.

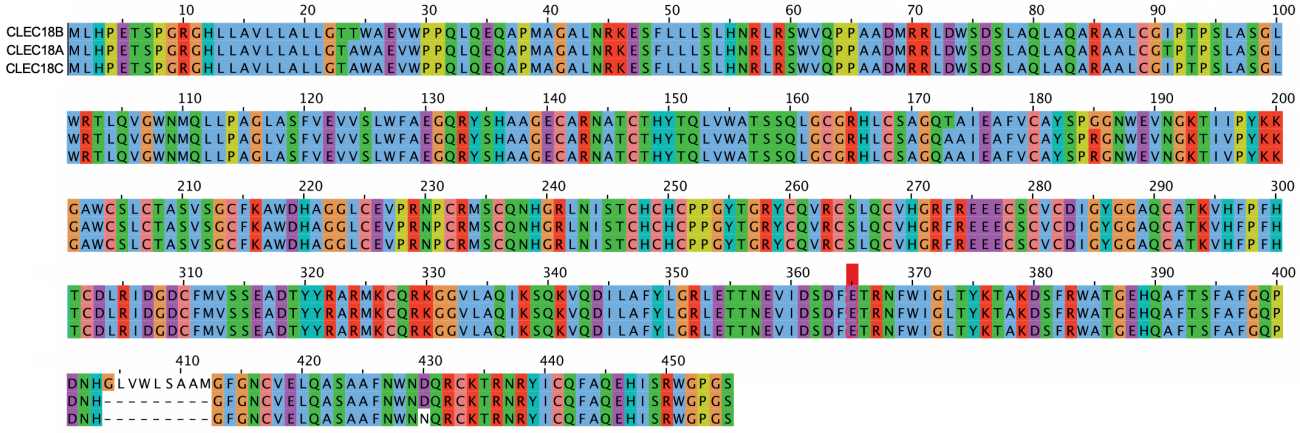
H. Tumor growth curve kinetics of *Clec18a*$^{+/+}$ (empty vector control) and *Clec18a*^{OE} breast cancer cell lines (E0771) in *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice. P values calculated assessed with two-way ANOVA with Šídák's multiple comparisons test. Abbreviations: OE = Overexpression.

I. Graphical abstract of the study.

Supplementary Figure 1.



B



Supplementary Figure 1. Conservation and expression of CLEC18A

A. Maximum Likelihood phylogenetic tree based on a multiple sequence alignment of selected CLEC18 protein sequences. Branches that are supported by SH-aLRT ≥ 80% and UFboot2 ≥ 95% are indicated by a grey dot.

B. Multiple sequence alignment of human CLEC18A, CLEC18B and CLEC18C proteins colored according to the Clustal scheme.

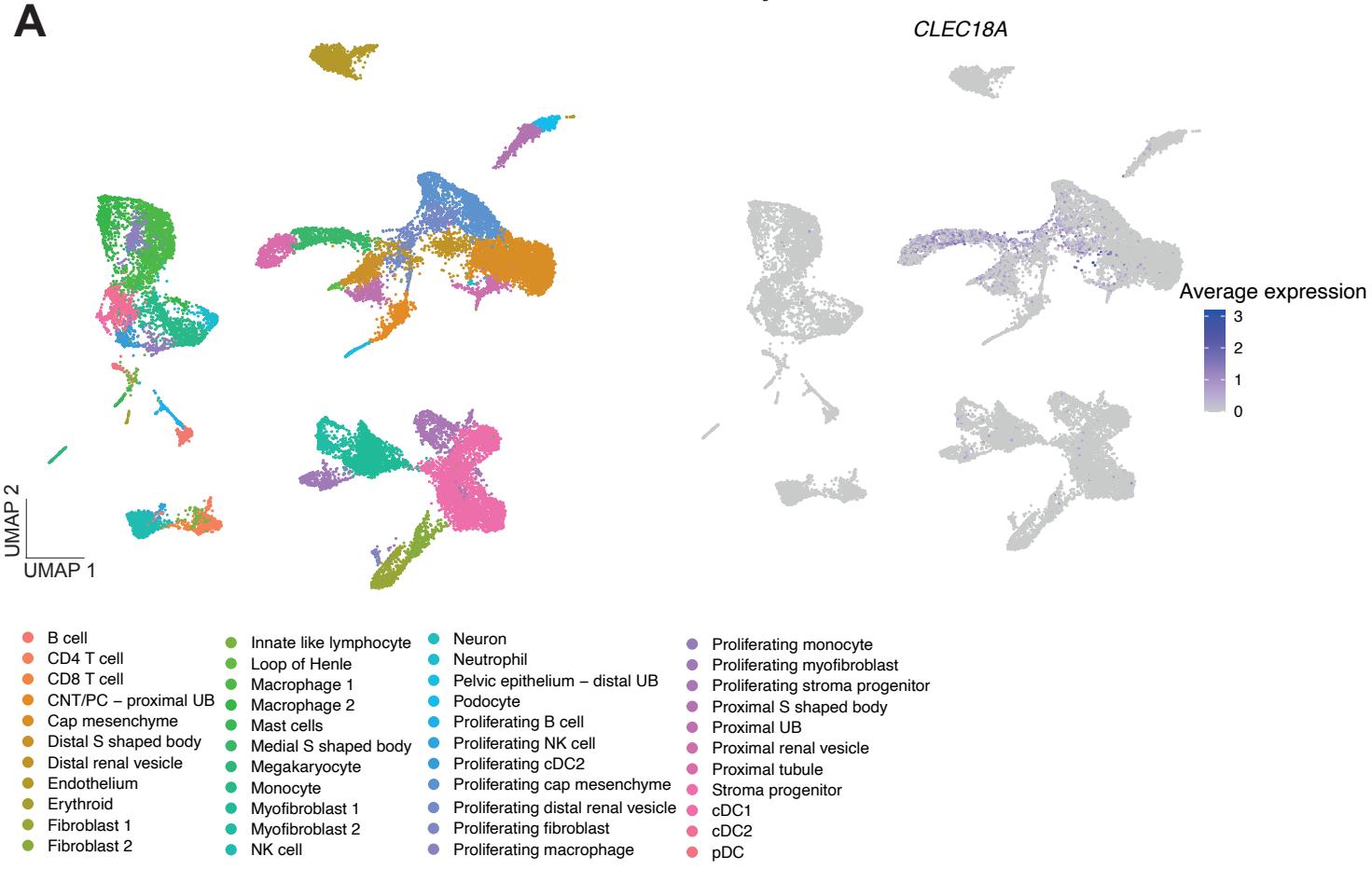
C. *Clec18a* expression in mouse organs determined by RT-qPCR. Organs were always harvested from there individual mice.

Supplementary Figure 2.

Stewart B.J. et al 2019

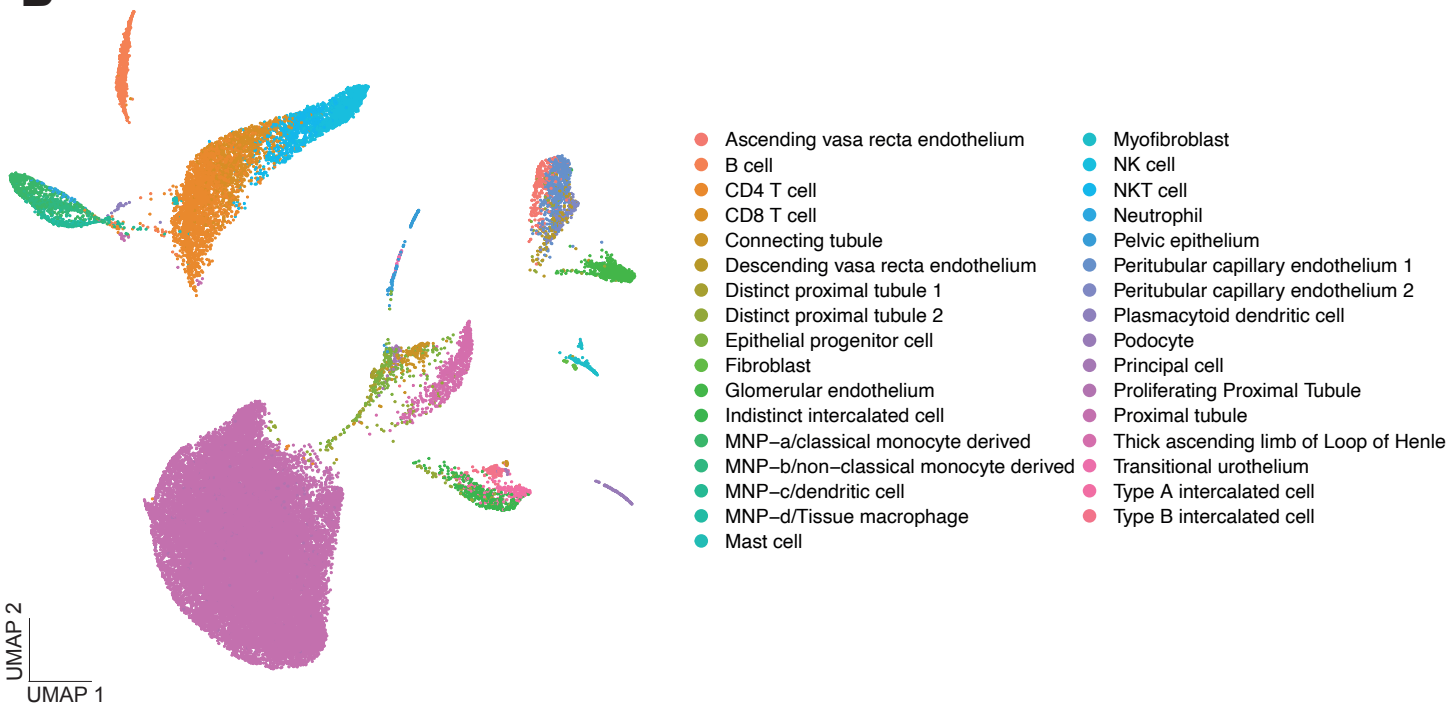
Fetal kidney

A



Adult kidney

B



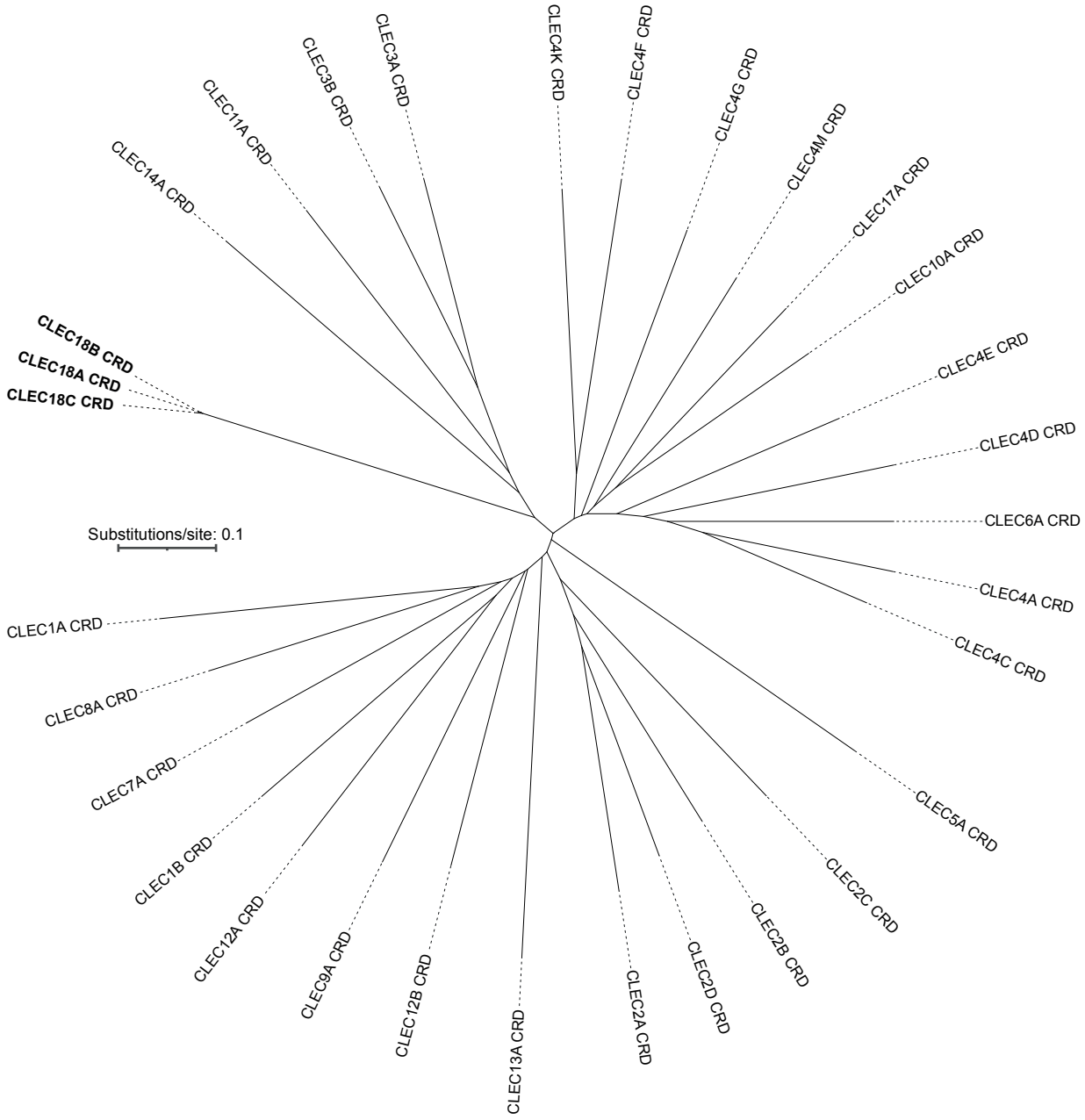
Supplementary Figure 2. scRNA-seq clustering of fetal kidney and adult kidney cells as accoridng to Stewart B.J. et al 2019 [18].

A. Re-analysis of scRNA-seq data from the kidney cell atlas [18] showing clustering of fetal kidney cells and *CLEC18A* expression in the full fetal kidney. Abbreviations: UMAP = Unifold manifold approximation and projection, CNT = Connecting tubule, PC = Principal cell, UB = Ureteric bud, NK cell = Natural killer cell, cDC1 = Conventional dendritic cell type 1, cDC2 = Conventional dendritic cell type 2, pDC = Plasmacytoid dendritic cell.

B. Re-analysis of scRNA-seq data from the kidney cell atlas [18] showing clustering of adult kidney cells. Abbreviations: UMAP = Unifold manifold approximation and projection, MNP = Mononuclear phagocyte, NK cell = Natural killer cell.

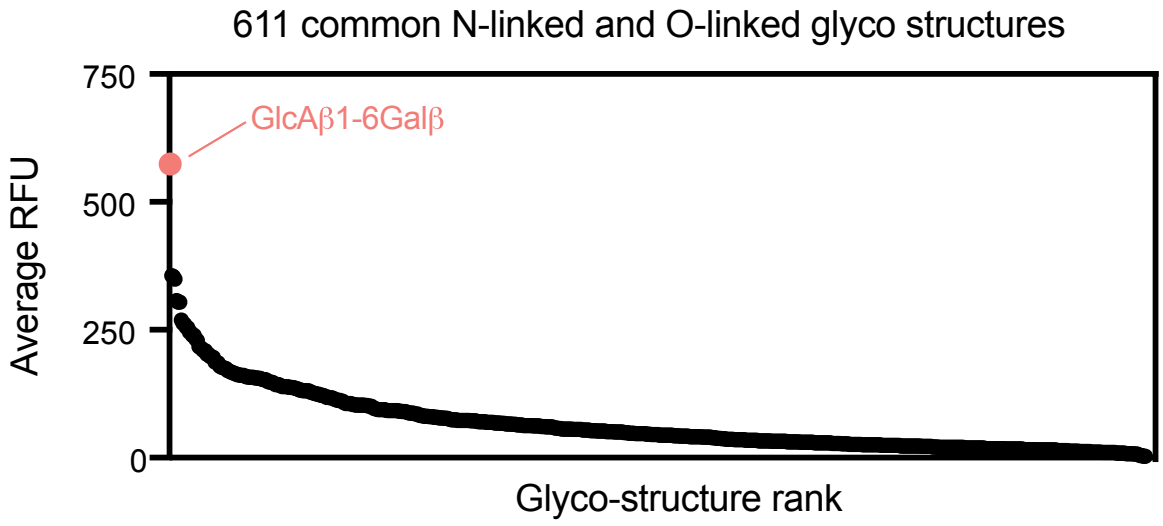
Supplementary Figure 3.

A



B

Huang YL et al. 2015

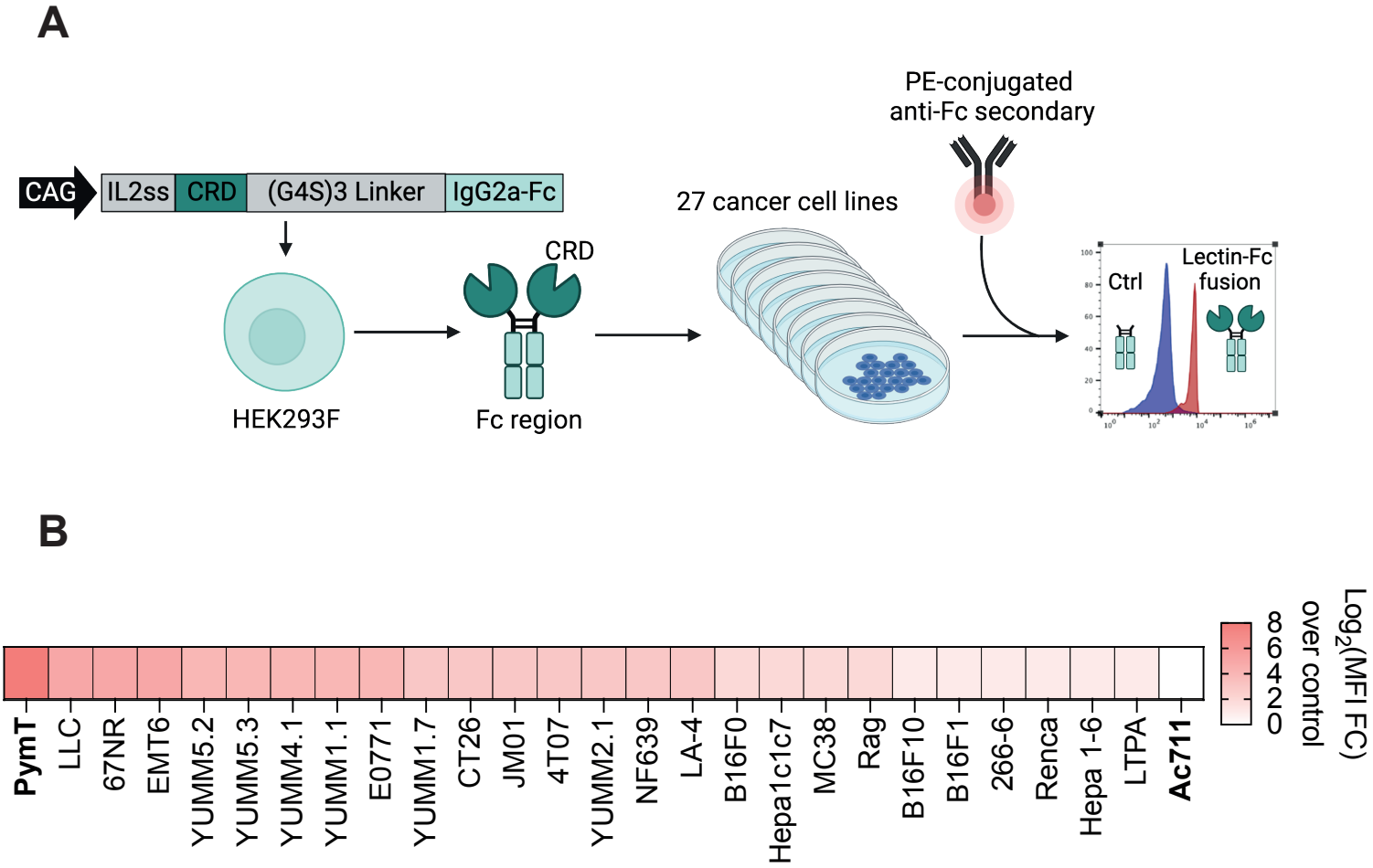


Supplementary Figure 3. CLEC18 CRDs have unique phylogeny and ligands

A. Unrooted neighbourhood joining phylogenetic tree based on the multiple sequence alignment of all included C-type lectin CRD sequences. Abbreviations: CRD = Carbohydrate recognition domain.

B. Re-analysis of a recombinant CLEC18A used on a GlycoArray containing 611 common N-linked and O-linked glyco structures [13].

Supplementary Figure 4.

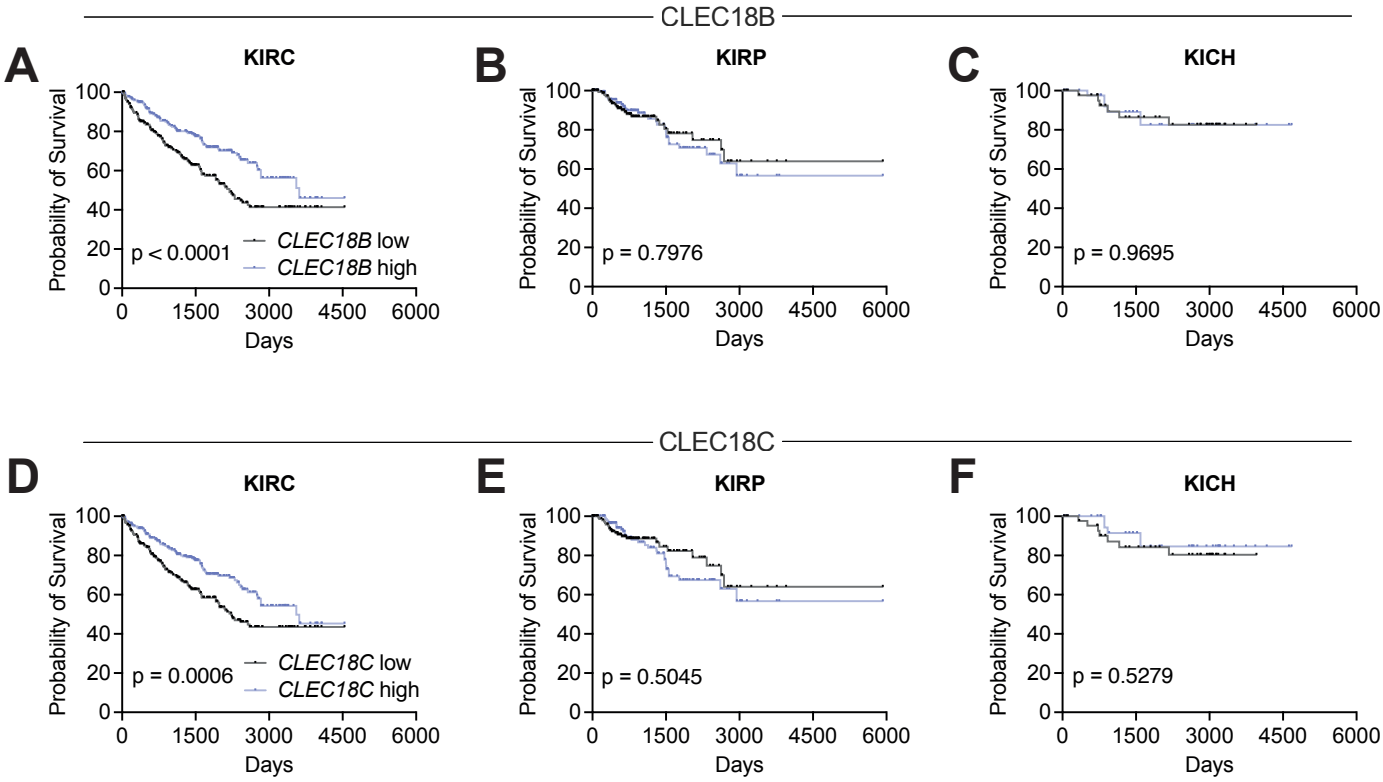


Supplementary Figure 4. CLEC18A-Fc binds cell lines with varying efficiency

A. Schematic of vector used for CLEC18A-Fc expression, subsequent expression and purification, and flow cytometry based readout used to identify cell lines CLEC18A-Fc binds to. Schematic created with BioRender.com.

B. Binding affinity of CLEC18A-Fc against a selection of cancer cell lines.

Supplementary Figure 5.



Supplementary Figure 5. CLEC18 exclusively promotes survival in ccRCC/KIRC

A. Kaplan meier survival curve for high vs low *CLEC18B* expression (high and low expression attributed based on the median) in KIRC. P value calculated with a Log rank test. Abbreviations: KIRC = Clear cell renal cell carcinoma.

B. Kaplan meier survival curve for high vs low *CLEC18B* expression (high and low expression attributed based on the median) in KIRP. P value calculated with a Log rank test. Abbreviations: KIRP = Papillary renal cell carcinoma.

C. Kaplan meier survival curve for high vs low *CLEC18B* expression (high and low expression attributed based on the median) in KICH. P value calculated with a Log rank test. Abbreviations: KICH = Chromophobe renal cell carcinoma.

D. Kaplan meier survival curve for high vs low *CLEC18C* expression (high and low expression attributed based on the median) in KIRC. P value calculated with a Log rank test. Abbreviations: KIRC = Clear cell renal cell carcinoma.

E. Kaplan meier survival curve for high vs low *CLEC18C* expression (high and low expression attributed based on the median) in KIRP. P value calculated with a Log rank test. Abbreviations: KIRP = Papillary renal cell carcinoma.

F. Kaplan meier survival curve for high vs low *CLEC18C* expression (high and low expression attributed based on the median) in KICH. P value calculated with a Log rank test. Abbreviations: KICH = Chromophobe renal cell carcinoma.

3.2 Publication 2: Urinary multi-omics reveal non-invasive diagnostic biomarkers in clear cell renal cell carcinoma

Authors:

Gustav Jonsson[1,2,3,*], Maura Hofmann[1,2,#], Tiago Oliveira[1,2,#], Ursula Lemberger[4,#], Karel Stejskal[1], Gabriela Kršťáková[1], Irma Sakic[1], Maria Novatchkova[1], Stefan Mereiter[1,2], Gerlinde Grabmann[5], Thomas Köcher[5], Zeljko Kikic[4], Gerald N. Rechberger[6], Thomas Züllig[6], Bernhard Englinger[4,$], Manuela Schmidinger[4,$], Josef M. Penninger[2,7,8,*]

Affiliations:

1. Institute of Molecular Biotechnology of the Austrian Academy of Sciences, IMBA, Dr. Bohr-Gasse 3, 1030 Vienna, Austria
2. Eric Kandel Institute, Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
3. Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, 1030, Vienna, Austria
4. Department of Urology, Center for Cancer Research and Comprehensive Cancer Center, Medical University of Vienna, Waehringer Guertel 18-20, 1090, Vienna, Austria
5. Vienna BioCenter Core Facilities (VBCF), 1030, Vienna, Austria
6. Institute of Molecular Biosciences, NAWI Graz, University of Graz, Humboldtstrasse 50, Graz, Austria
7. Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Canada
8. Helmholtz Centre for Infection Research, Braunschweig, Germany

# These authors contributed equally to this study as second authors

$ These authors contributed equally to this study as second-last authors

* Correspondence to: Josef M. Penninger (josef.penninger@helmholtz-hzi.de) and Gustav Jonsson (gustav.jonsson@imba.oeaw.ac.at)

Author contributions:

G.J. and J.M.P conceived the study. G.J. performed and designed all experiments with input and help from all co-authors as follows: M.H. with mass spectrometry sample preparation and

overall data analysis; T.O. with mass spectrometry; U.L, Z.K, B.E. and M.S. with ethical approval and patient sample collection; S.M. with data analysis and interpretation; M.N. with bioinformatics; G.G. and T.K. with metabolomics; G.K. and K.S. with PRM-MS; G.N.R and T.Z. with lipidomics. G.J. and J.M.P wrote the paper with input from all authors.

Urinary multi-omics reveal non-invasive diagnostic biomarkers in clear cell renal cell carcinoma

Gustav Jonsson[1,2,3,*], Maura Hofmann[1,2,#], Tiago Oliveira[1,2,#], Ursula Lemberger[4,#], Karel Stejskal[1], Gabriela Kršsáková[1], Irma Sakic[1], Maria Novatchkova[1], Stefan Mereiter[1,2], Gerlinde Grabmann[5], Thomas Köcher[5], Zeljko Kikic[4], Gerald N. Rechberger[6], Thomas Züllig[6], Bernhard Englinger[4,$], Manuela Schmidinger[4,$], Josef M. Penninger[2,7,8,*]

1. Institute of Molecular Biotechnology of the Austrian Academy of Sciences, IMBA, Dr. Bohr-Gasse 3, 1030 Vienna, Austria
2. Eric Kandel Institute, Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
3. Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, 1030, Vienna, Austria
4. Department of Urology, Center for Cancer Research and Comprehensive Cancer Center, Medical University of Vienna, Waehringerguertel 18-20, 1090, Vienna, Austria
5. Vienna BioCenter Core Facilities (VBCF), 1030, Vienna, Austria
6. Institute of Molecular Biosciences, NAWI Graz, University of Graz, Humboldtstrasse 50, Graz, Austria
7. Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Canada
8. Helmholtz Centre for Infection Research, Braunschweig, Germany

# These authors contributed equally to this study as second authors
$ These authors contributed equally to this study as second-last authors
* Correspondence to: Josef M. Penninger (josef.penninger@helmholtz-hzi.de) and Gustav Jonsson (gustav.jonsson@imba.oeaw.ac.at)

Abstract

Clear cell renal cell carcinoma (ccRCC) is the kidney malignancy with the highest incidence and mortality rates. Despite the high patient burden, there are no biomarkers for rapid diagnosis and public health surveillance. Urine would be an ideal source of ccRCC biomarkers due to the low invasiveness, easy accessibility, and the kidney's intrinsic role in filtering urine. In the present work, by combining proteomics, lipidomics and metabolomics, we detected urogenital metabolic dysregulation in ccRCC patients with increased lipid metabolism, altered mitochondrial respiration signatures and increased urinary lipid content. Importantly, we identify three early-stage diagnostic biomarkers for ccRCC in urine samples: Serum amyloid A1 (SAA1), Haptoglobin (HP) and Lipocalin 15 (LCN15). We further implemented a parallel reaction monitoring mass spectrometry protocol for rapid and sensitive detection of SAA1, HP and LCN15 and combined all three proteins into a diagnostic UrineScore. In our discovery cohort, this score had a performance accuracy of 96% in receiver operating characteristic curve (ROC) analysis for classification of ccRCC versus control cases. Our data identifies tractable and highly efficacious urinary biomarkers for ccRCC diagnosis and serve as a first step towards the development of more rapid and accessible urinary diagnostic platforms.

Introduction

Renal cell carcinomas (RCCs) are a heterogenous group of cancers arising from the proximal convoluted tubule of the kidney. Many histological subtypes of renal cell carcinomas have been described, with the three most common ones being clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC) and chromophobe renal cell carcinoma (chRCC). Out of these three subtypes, ccRCC is the most common, making up approximately 80% of all RCC cases, and also the deadliest [1, 2]. ccRCC is primarily driven by loss-of-function mutations or epigenetic silencing of the *Von Hippel Lindau (VHL)* tumor suppressor, leading to the initiation of a genetic hypoxia program which drives tumor progression [3, 4]. Despite this clearly identified mechanism, solely screening for *VHL* mutations for disease detection is not sufficient since not all patients present with *VHL* gene mutations. Furthermore, multiple additional co-drivers have also been identified, such as PBRM1, SETD2, KDM5C, or BAP1 [5, 6], convoluting genetic-based disease screening.

Despite significant advances in the treatment of ccRCC using small molecule inhibitors and immunotherapy [7-9], early screenable markers for diagnosis are still lacking and most patients are diagnosed through computed tomography when a tumor is already suspected, for example due to palpable renal masses or hematuria [10]. Hematuria is a very common sign of ccRCC, and other renal cell malignancies, but in and of itself insufficient for a reliable diagnosis. The proximal convoluted tubule, where ccRCC arises, is responsible for secreting and reabsorbing solutes between blood and urine [11] increasing the likelihood that secreted or shed biomarkers are present and detectable in the urine. Furthermore, ccRCC predominantly affects people of the age of 60 and upwards, and very rarely younger individuals [10]. In many countries, the older population is already part of population-based screening programs and health check-ups in which urine is routinely donated and analyzed, making urine samples an ideal source for early detection of ccRCC. Urine also allows easy longitudinal sampling and is minimally invase. Omics approaches have been used on urine samples from renal cell carcinomas in the past for the purpose of biomarker discovery, predominantly metabolomics due to the natural abundance of high levels of excreted metabolites as waste products in urine [12]. However, multiple omics approaches have yet to be integrated on the same urine samples to draw multi-modal conclusions about the dysregulation of the urinary landscape in ccRCC, and how this can be exploited for biomarker discovery.

In the present work, by combining proteomics, lipidomics and metabolomics, we detected urogenital metabolic dysregulation in ccRCC patients with increased lipid metabolism, mitochondrial respiration signatures and increased urinary lipid content. We further aimed to explore urinary biomarkers to reliably detect ccRCC. In a clinical cohort of controls, ccRCC patients and non-clear renal cell carcinoma (nccRCC) patients, we discovered and validated three proteins as ccRCC diagnostic biomarkers: SAA1, HP, and LCN15. We further developed a parallel reaction monitoring mass spectrometry (PRM-MS) signature for rapid and sensitive quantitative detection of these three marker proteins in patient urine, discriminating between controls and ccRCC patients with a performance accuracy of 96%.

Results

Urine from ccRCC patients is indicative of metabolic dysregulation

To characterize the landscape of ccRCC urine and to potentially derive biomarkers for the disease we initially performed proteomics on urine sediment and supernatant on a cohort of 22 ccRCC patients and 12 controls **(Fig. 1A, Supplementary Table 1**). Of note, in pilot studies we found that precipitating proteins in two steps from the urine samples (tested on supernatant) starting with an acetone precipitation followed by a chloroform:methanol precipitation yielded the highest number of unique peptides and proteins identified (**Supplementary Fig. 1**). This method was subsequently used for all protein preparations. All discovery-phase proteomics analysis was done through data independent acquisition (DIA). Gene ontology (GO) term analysis of all upregulated proteins in ccRCC urine supernatants from proteomics, compared to controls, predominantly revealed metabolic dysregulation affecting lipid metabolism and function through, e.g. fatty acid transport and high-density lipoparticle remodeling (**Fig. 1B, Supplementary Table 2**). Further, GO term analysis on sediment proteomics from control and ccRCC patient urine samples showed that mitochondrial respiration and electron transport processes were upregulated in ccRCC patients (**Fig. 1C, Supplementary Table 3**).

Since the proteomics data indicated dysregulated metabolism throughout the urogenital tract, we further performed metabolomics and lipidomics on the urine samples. Untargeted metabolomics predominantly revealed downregulated metabolites in ccRCC patients (**Supplementary Fig. 2A**). After filtering the hits further to only include metabolites from the mzCloud database and our in-house validated database, 36 downregulated metabolites remained (**Supplementary Fig 2A-B, Supplementary Table 4**); some examples of which are shown in **Supplementary Fig. 2C**. GO term analysis on the 36 downregulated metabolites,

matched against the KEGG and SMPDB databases, revealed downregulation of (i) aspartate, alanine and glutamate metabolism, (ii) glyoxylate metabolism, (iii) glutathione metabolism and (iv) carnitine synthesis (**Fig. 1D**).

Our lipidomics analysis revealed that 3 out of 9 detected lipid classes were upregulated in ccRCC compared to controls: Coenzyme Q10, phosphatidylcholines (PCs) and phosphatidylethanolamines (PEs) (**Fig. 1E-F, Supplementary Table 5**). Combined, most of the measured individual PC and PE species were significantly upregulated in ccRCC compared to controls (**Supplementary Fig. 3A-B**). In summary, the proteomics, metabolomics and lipidomics data indicate dysregulated metabolism with increased electron transport chain activity and increased lipid metabolism throughout the urogenital tract, summarized in **Fig. 1G**. The observed lipid phenotype is likely linked to cancer cells whereas electron transport chain phenotypes are linked to healthy, shed epithelium in the urine sediment. ccRCC cells are known to be lipid laden and accumulate high levels of lipids. Increased PC and PE levels in the urine likely come from increased lipid transport in these cells. Further, carnitine is important for transport of fatty acids through the mitochondrial membrane, leading to their breakdown in β-oxidation, meaning that the reduction of carnitine is likely to contribute to the increased lipid content in the cells. Combined, the lipid landscape of the urine supernatant appears to be representative of what is going on in the tumors. ccRCC is known to be independent of oxidative phosphorylation [13] indicating that upregulated respiratory chain pathways in the urine sediment most likely come from increased shedding of healthy epithelium or other shed urogenital cells, and not actually from the tumor itself. We further detected upregulation of CoQ10 which is important for electron chain transport function, and downregulation of alternative entry points into the TCA cycle (alanine, aspartate and glutamate metabolism, and glyoxylate metabolism) (**Fig. 1E-G),** indicating that conventional glycolysis is used for increased oxidative phosphorylation.

Lipidomics and metabolomics provide putative diagnostic biomarkers for ccRCC

ccRCC is the deadliest type of renal carcinoma (**Fig. 2A**) with a tendency to appear asymptomatic at early stages leading to complications at the later stages of disease. Compared to many other malignancies, ccRCC lacks liquid biopsy biomarkers for early diagnosis and prognosis [14, 15]. We hypothesized that our multi-omics dataset could lead to the identification of reliable early-stage biomarkers since a majority of our ccRCC cohort, included in the analyses, consisted of early pT1 stage tumors (**Fig. 2B**). Initially we assessed the ability

of the three significantly upregulated lipid classes (CoQ10, PC and PE) to distinguish between ccRCC and controls in an area under the receiver operating characteristic curve (AUROC) analysis (**Fig. 2C**). CoQ10, PC and PE all had a good performance of AUROC = 0.7992, AUROC = 0.8826 and AUROC = 0.8258, respectively, in distinguishing between ccRCC and control cases. Furthermore, when we assessed all significantly upregulated individual lipid species, all of them had an AUROC > 0.7, with PC 38:3 and PC 38:6 having the highest AUROC of 0.9129 and 0.8996, respectively (**Fig. 2D**), indicating the potential use of PC family members in ccRCC diagnosis.

Similar trends were observed when assessing downregulated metabolites. All significantly downregulated metabolites presented with an AUROC > 0.7 in being able to distinguish between ccRCC and controls, with citraconic acid, citric acid and DL-Isocitric acid scoring the highest AUROC values of 0.9545, 0.9432, and 0.9356, respectively (**Fig. 2E**). These data indicate that altered lipid and metabolite signatures could be used to distinguish ccRCC from healthy controls in urine.

Proteomics provides high-confidence diagnostic biomarkers for ccRCC

Despite the statistical validity of lipids and metabolites in distinguishing between ccRCC and control cases and generating interesting biological data, for biomarker utilization they do come with a set of challenges. For lipidomics, it is difficult to achieve complete analytical coverage and accurate abundancy quantification remains an issue [16]. Furthermore, in our study the total number of identified lipid species was on the lower side and all species were not identified in every single sample. For metabolomics, we only detected significantly downregulated metabolites in ccRCC patients. Urine as a sample has low levels of analytes, and basing diagnosis on downregulation of already lowly abundant analytes is challenging.

For these reasons, we focused our efforts at finding an easily assessable biomarker on proteomics. Mass spectrometry-based proteomics are frequently available and established in hospital diagnostic laboratories compared to lipidomics and metabolomics. Protein level diagnosis further offers the benefit of being able to develop ELISA based approaches for cost-effective and routine disease screening. Our bulk proteomics data on both urine sediment and supernatant revealed upregulated proteins with potential as diagnostic biomarkers; e.g. NAT10, APOL1, NDUFS7 and MRPL48 for sediment (**Fig. 3A**) and SAA1, HP and LCN15 for supernatant (**Fig. 3B**). Mass spectrometry data is available in **Supplementary Table 2-3**. For

any potential diagnostic biomarker, in addition to being able to distinguish between ccRCC and controls, we also wanted to test if such markers can distinguish between ccRCC and non-clear cell renal cell carcinomas (nccRCC), such as pRCC and chRCC. For this purpose, we extended our cohort to also include 9 nccRCC patients, from which 8 patients had pRCC and 1 patient chRCC (**Fig. 3C, Supplementary Table 1**). ccRCC patients had the highest amounts of protein in the urine supernatant (**Supplementary Fig. 4A**). The overall amount of protein found in the urine of ccRCC patients was independent of the tumor stage (**Supplementary Fig. 4B)**. Furthermore, for leukocytes and inflammatory damage markers (Creatinine and C-reactive protein) in the serum, no differences were found between ccRCC and nccRCC patients, nor between different stages of ccRCC (**Supplementary Fig. 4C, D**).

When comparing nccRCC urine sediments to controls, NAT10 and MRPL48 were the highest upregulated proteins, just as for ccRCC urine sediments compared to controls (**Fig. 3D, Supplementary Fig. 5A**). Overall, the upregulated protein profiles were similar for both ccRCC and nccRCC, with LRCH4 and TP53I3 being the only proteins with the potential of distinguishing between ccRCC and nccRCC (**Fig 3D, Supplementary Fig. 5B**). For supernatants however, different proteins were upregulated between cancer and control for ccRCC and nccRCC, except for MATR3 (**Fig. 3E, Supplementary Fig. 5C-D**). Due to the different urinary protein profiles of ccRCC and nccRCC, urine supernatant proteins were chosen as the preferred source of diagnostic biomarkers. Furthermore, reasonable protein quantities for proteomics could not be extracted from all samples (**Supplementary Table 3**). Serum Amyloid A1 (SAA1), Haptoglobin (HP) and Lipocalin 15 (LCN15) were selected as the putative biomarkers based on the following criteria: (i) having a $\log_2$(Fold change) > 2 in ccRCC compared to controls, (ii) upregulated in the ccRCC and control comparison, and not the nccRCC and control comparison which eliminates MATR3 and (iii) not being male or female specific, e.g. CRISP1 is male specific.

Parallel reaction monitoring mass spectrometry for rapid detection of SAA1, HP and LCN15 in ccRCC urine

Conventional bulk proteomics allows discovery at the full proteome level but is not suitable for rapid diagnostics and population screenings in a clinical setting due to long instrument run times, costs, and computationally intensive data analysis. Therefore, we utilized parallel reaction monitoring mass spectrometry (PRM-MS) to develop a tractable diagnostic modality for SAA1, HP and LCN15. PRM-MS is an ion monitoring technique allowing parallel high-

resolution detection of peptides of interest, drastically reducing the run time per sample and increasing specificity compared to bulk proteomics and other ion monitoring techniques [17].

In our PRM-MS approach, we included all peptides for SAA1, HP and LCN15 detected by our bulk proteomics approach with clearly defined elution patterns (**Supplementary Table 6**). For standardization, we included peptides from three normalization proteins, namely Uromodulin (UMOD), Kallikrein-1 (KLK1) and Apolipoprotein D (APOD). A normalization protein was defined as a protein which was found in every sample in all three groups (Control, nccRCC and ccRCC) and with similar expression levels across all three experiment groups per protein (**Supplementary Table 2**). The peptide area was calculated for every selected peptide and the PRM score for each diagnostic protein was determined by dividing the sum of all peptide areas from individual proteins of interest (SAA1, HP, LCN15) with the sum of all peptides from all three normalization proteins (**Fig. 4A**).

The PRM scores for SAA1, HP and LCN15 were all significantly higher in ccRCC patients compared to controls (**Fig. 4B**). The classification performance of each PRM score was calculated by the AUROC method and was evaluated to be 0.88, 0.89 and 0.84 for SAA1, HP and LCN15, respectively (**Supplementary Fig. 6**).

Lastly, all three individual PRM scores were combined into a cumulative UrineScore, detailed in the Methods section. In brief, the 95% confidence interval (CI) of the median was calculated for all PRM parameters for the control samples. Subsequently, each sample from the control, nccRCC and ccRCC groups was attributed a score of 1 per protein (SAA1, HP and LCN15) which is a value higher than the upper limit of the control 95% CI, meaning that the UrineScore is an integer between 0-3 per control or patient. This UrineScore was the highest for the ccRCC group (**Fig. 4C**) and had a very high-performance accuracy in distinguishing between control and ccRCC samples (AUROCC = 0.96), and also performed well between nccRCC and ccRCC samples (AUROCC = 0.79) (**Fig. 4D**). These data indicate that PRM-MS allows more rapid and sensitive diagnostic power compared to bulk proteomics, and that SAA1, HP and LCN15 can be used to diagnose ccRCC in our discovery cohort.

Discussion

In 2020, approximately 431000 new cases of renal cell carcinomas were diagnosed worldwide [18]. Out of all of these cases, it is estimated that 80% of the renal cell carcinomas are ccRCC.

With a growing incidence rate, there is increased interest in cost-effective and sensitive population-wide screening programs. Despite the high clinical interest, there are no validated biomarkers for clinical screening, and common diagnostic procedures such as computed tomography are not suitable for population-based screenings programs [19, 20].

Utilizing urinary proteomics, metabolomics and lipidomics, we uncovered aberrant metabolism throughout the urogenital tract of ccRCC patients. Our data suggests increased mitochondrial respiration and lipid metabolism. ccRCC is known to have a lipid rich cytoplasm [21]. We discovered increased lipid transport and a reduction in carnitine synthesis. Although our data comes from the urine and not directly from the tumor, one potential explanation for the increased lipid accumulation in ccRCC tumors could be a reduction of carnitine, which is essential for the breakdown of free fatty acids and potentially lipid transport to the tumor microenvironment, and not only increased lipid transport through the reverse cholesterol pathway in which cholesterol is recycled to the liver [22]. Amongst the individual lipid species detected, we noticed an increase in certain PE and PC species indicative of increased tissue damage and membrane shedding into the excreted compartment. We further found increased levels of CoQ10, corroborating our observations of increased mitochondrial respiration. Generally, ccRCC tend to be independent of oxidative phosphorylation as a source of energy [13], indicating that the mitochondrial respiration phenotypes we encountered in urine sediment are likely coming from shed epithelium and surrounding tissue damage from the tumor.

Proteomics on the urine supernatants proved the most reliable and feasible for diagnostic biomarker discovery. Bulk proteomics uncovered three putative diagnostic biomarkers (SAA1, HP and LCN15) in a small clinical cohort. Furthermore, through the establishment of PRM-MS for rapid and sensitive detection of these three proteins of interest we could derive PRM scores for SAA1, HP and LCN15 with high diagnostic power. When these PRM scores were combined into an overarching UrineScore, we calculated an overall performance in an AUROC analysis of 96% in distinguishing between healthy controls and ccRCC. The UrineScore also performed well in distinguishing between ccRCC and nccRCC patients, though this needs to be confirmed in larger patient cohorts.

The cellular source of SAA1, HP and LCN15 and how these three proteins are being filtered/shed into the urine remains to be answered. All three proteins are secreted proteins,

meaning that they could directly be produced at high levels in the tumors and subsequently secreted into the urine. However, SAA1 and HP are also produced as acute phase proteins in the liver, and LCN15 is produced in the gastrointestinal tract. Future studies will have to show how large of a proportion of the markers found in the urine are directly derived from the tumor and/or other other cellular sources. It is possible that circulating serum proteins are found at higher levels in the urine of ccRCC patients due to tumor-induced tissue damage of the filtration unit of the kidney, effectively increasing the leakiness of serum proteins into the urine which is found in ccRCC associated hematuria. However, this does not appear to be a general mechanism since we in that case would likely have found a larger range of serum proteins as diagnostic markers.

The urine proteome of ccRCC patients has been studied before. For example, differences in urine protein content has been found amongst ccRCC patients depending on disease prognosis [23, 24] and venous infiltration [25]. Furthermore, attempts have been made at identifying diagnostic and prognostic protein biomarkers in the urine of ccRCC patients [26, 27]. However, none of these studies identified SAA1, HP or LCN15 in their analysis. Lastly, our three proteins of interest also differentiate between ccRCC and nccRCC, highlighting the specificity of these markers towards ccRCC. Validation trials need to be conducted to validate this readily screenable UrineScore in larger, independent clinical cohorts but the UrineScore highlights the potential of urine based diagnostic scoring not only for identifying renal cell carcinomas, but also for distinguishing between different histological subtypes.

Materials and Methods

Patient cohort

The study was approved by the ethical commission at the Medical University of Vienna (Ethik-Kommission Medizinische Universität Wien), study number 2224/2021, project title: Urinproteomik zur Validierung von Biomarkern für das klarzellige Nierenkarzinom – Pilotstudie (UrineProt).

Urine was collected from 40 patients who presented with a suspected primary renal mass at the Department of Urology of the Medical University of Vienna. Out of the 40 patients, 9 were excluded due to the renal mass being identified as something other than a renal cell carcinoma, such as oncocytomas, cysts, angiomyolipoma, papillary adenoma or a kidney-lodged metastasis. From the remaining 31 patients, 22 patients were characterized as ccRCC, 8 as pRCC and 1 as chromophobe RCC through histological assessment by a trained pathologist. For further analysis, the pRCC and chromophobe RCC patients were combined into one group of non-clear cell renal cell carcinoma (nccRCC) patients. Leukocytes, Creatinine and CRP levels were obtained in the routine blood analysis using standard methods at the Institute of Laboratory Medicine, Vienna General Hospital.

Protein precipitation

3 ml of ccRCC patient, nccRCC patient or control urine were centrifuged at 1000xg for 10 min. The urine supernatant was isolated and mixed with 12 ml of acetone and incubated for a minimum of 2 hours at -20^{o}C to allow for protein precipitation. Samples were centrifuged at 3000xg for 60 min and the supernatant discarded. The protein pellet was then resuspended in 100 µl of 8M Urea and a second precipitation step was performed to increase protein yields and to discard residual contaminants using chloroform and methanol [28]. 400 µl methanol was added to the protein solution and vortexed extensively. Subsequently, 200 µl of chloroform and 300 µl of water was added with vortexing steps in between. The mixtures were centrifuged at 10000xg for 15 min at room temperature to ensure phase separation. Following centrifugation, the top aqueous layer was removed leaving the lower chloroform fraction and white protein interface. To the remaining mixture, 400 µl of methanol was added followed by vortexing and centrifugation at 10000xg for 5 min. Supernatant was removed and the methanol wash was repeated twice. After the last methanol wash, the pellet was dried in a speed vacuum and resuspended in 8M Urea.

Enzymatic digestion of precipitated proteins

Precipitated proteins were reduced in 10 mM DTT (Roche) at 37°C for 1 hour and diluted to 4M Urea. Following reduction, proteins were subsequently alkylated with 20 mM Indole-3-acetic acid (Merck) for 1h at room temperature in the dark and diluted to 2M Urea. Following reduction and alkylation, proteins were initially enzymatically digested with LysC (Lysyl Endopeptidase®, Mass Spectrometry Grade, FUJIFILM) at 37°C for 2hrs according to the manufacturer's instructions. Lastly, peptides were further digested with trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega) overnight at 37°C according to the manufacturer's instructions.

Bulk proteomics

Data-independent acquisition mass spectrometry. Individual peptide samples were analyzed by LC–MS/MS. The nano HPLC system used was an UltiMate 3000 nano HPLC RSLC (Thermo Scientific) equipped nano-electrospray source (CaptiveSpray source, Bruker Daltonics), coupled to a timsTOF HT mass spectrometer (Bruker Daltonics). Peptides samples were injected on a pre-column (PepMap C18, 5 mm × 300 µm × 5 µm, 100 Å pore size, Thermo Scientific) with 2% ACN/water (v/v) containing 0.1% TFA at a flow rate of 10 µL/min for 10 min. Peptides were then separated on a 25 cm Aurora ULTIMATE series HPLC column equipped with an emitter (CSI, 25 cm × 75 µm ID, 1.7 µm C18, IonOpticks) operating at 50°C and controlled by the Column Oven PRSO-V1-BR (Sonation), using UltiMate 3000 (Thermo Scientific Dionex). The analytical column flow was run at 300 nL/min with two mobile phases: water with 0.1% FA (A) and water with 80% acetonitrile and 0.08% formic acid (B). A and B were applied in linear gradients as follows (only B percentages reported): starting from 2% B: 2%-10% B in 10 min, 10%-24% B in 35 min, 24%-35% B in 15 min, 35%-95% B in 1 min, 95% for 5 min, and finally the column was equilibrated in 2% B for next 10 min (all % values are v/v; Water and ACN solvents were purchased from Thermo Fisher Scientific at LC-MS grade).

The LC system was coupled to a TIMS quadrupole time-of-flight mass spectromecter (timsTOF HT, Bruker Daltonics) and samples were measured in dia-PASEF mode. The CaptiveSpray source parameters were: 1600 V capillary voltage, 3.0 l/min dry gas, and 180 °C dry temperature. MS data was acquired in the MS scan mode, using positive polarity, 100-1700 m/z range, mobility range was set up from 0.64-1.42 V s/cm^2, ramp time was set to 166 ms and

the estimated cycle time was 1.52s. Collision energy was 20 eV at $1/K_0$ 0.6 V s/cm^2, and 80 eV at $1/K_0$ 1.6 V s/cm^2. Automatic calibration of ion mobility was enabled. The timsTOF HT was operated in DIA mode where 1 MS1 scan was followed by 8 DIA-PASEF frames.

Proteomics data analysis. DIA data was analyzed in Spectronaut 18.5 [29] (Biognosys). Trypsin/P was specified as a proteolytic enzyme and up to 2 missed cleavages were allowed in the Pulsar direct DIA search. Dynamic mass tolerance was applied for the TOF calibration. Peptides were matched against the human UniProt database (20230710, 20 586 sequences), with common contaminants (344 sequences) and common tags (28 sequences) appended. Carbamidomethylation of cysteine was searched as fixed modification, whereas oxidation of methionine and acetylation at protein N-termini were defined as variable modifications. Peptides with a length between 7 and 52 amino acids were considered and results were filtered for 1% FDR at the peptide spectrum match (PSM), Peptide and Protein Group Level. Quantification was performed as specified in Biognosys BGS Factory Default settings, grouping Peptides by Stripped Sequence, and performing protein inference using IDPicker. For normalization Cross-Run Normalization in Spectronaut was activated.

Spectronaut results were exported using Pivot Reports on the Protein and Peptide level and converted to Microsoft Excel files using our in-house software MS2Go. For DIA data MS2Go utilizes the python library msReport (developed at the Max Perutz Labs Proteomics Facility) for data processing. Missing values were imputed with values obtained from a log-normal distribution with a mean of 30 in msReport and statistical significance of differentially expressed proteins was determined using limma-moderated Benjamini–Hochberg-corrected two-sided t-test [30].

Metabolomics

Metabolites were extracted from each sample by mixing 20 µl of urine supernatants with 200 µl methanol. Samples were subsequently dried down in a vacuum centrifuge and resuspended in 0.1% formic acid. Creatinine levels were determined in a targeted LC-MS/MS experiment. Normalized to the amount of creatinine determined, another aliquot of each extracted sample was evaporated and resuspended in 130 µl ACN:H2O (80:20). Samples were then centrifuged at 4°C for 10 min at 16000 g and transferred to a glass HPLC vial. 2 µl of all samples were pooled and used as a quality control (QC) sample. Samples were randomly assigned into the autosampler, and metabolites were separated on an iHILIC®-(P) Classic HPLC column

(HILICON AB, 100 x 2.1 mm; 5 µm; 200 Å, Sweden) with a flow rate of 100 µl/min delivered through an Ultimate 3000 HPLC system (Thermo Fisher Scientific, Germany). The stepwise gradient started at 90% A (ACN) and took 21 min to 60% B (25 mM ammonium bicarbonate) followed by 5 min hold at 80% B and subsequent equilibration phase at 90% A with a total run time of 35 min. Sample spectra were acquired by a high-resolution tandem mass spectrometer (Q-Exactive Focus, Thermo Fisher Scientific, Germany) in full MS mode. Metabolites were ionized via electrospray ionization in polarity switching mode after HILIC separation. Ionization potential was set to +3.5/-3.0 kV, the sheet gas flow was set to 20, and an auxiliary gas flow of 5 was used. Samples were subjected to randomized analysis, flanked by a blank and a QC sample for background correction and data normalization, respectively, occurring after every set of 8 samples. QC samples were additionally measured in data-dependent and confirmation mode to obtain MS/MS spectra for identification. The obtained data set was processed by "Compound Discoverer 3.3 SP2" (Thermo Fisher Scientific). Compounds were annotated through searching against our internal mass list database which was generated with authentic standard solutions. Additional compound annotation was conducted by searching the mzCloud database.

Lipidomics

Lipids were extracted according to the Matyash protocol [31] using 3 ml of Urinary supernatant. Internal standard mix (PE 34:0, 830456P; PS 34:0, 840028P; LPC 17:1, 855677C; SM d35:1, 860585; purchased form Avanti Polar Lipids, USA, and PC 34:0, 37-1700-7; TG 54:0, 33-1835-9; purchased from Larodan, Sweden) was added to the samples before extraction. The organic phase of the final extraction was dried in a vacuum centrifuge and resolved in 500 µl 2-propanol:MeOH:H_2O (70:25:10, v:v:v) before injection.

Samples were analyzed with reversed phase-UHPLC (BEH-C18, 2.1x 150 mm, 1.7 µm, Waters, Milford, USA) QTOF-MS (1290 Infinity II and 6560 IM-QTOF-MS, Agilent, Waldbronn, Germany) in positive/negative ESI QTOF-only mode. For the gradient elution, an aqueous eluent A and a 2-propanol eluent B were used, both with the following additives: Ammonium acetate (10 mM), phosphoric acid (8 µM), and formic acid (0.1 vol%). The gradient started with 60% eluent A for 0.5 min, followed by a linear decrease over 8.5 min to 20% and within 13 min to 0% A. This composition was held constant for 2.5 min and then returned to initial conditions for 5 min prior to the next injection. Eluent flow was constant 150 µl/min. In positive mode 1 µl and in negative ion mode 5 µl were injected. Column temperature was 50°C. The ESI instrument parameters were in positive mode: Gas temp 300°C, flow 10

l/min, Nebulizer 50 psi, sheath gas temp 400°C, flow 12 l/min), and in negative mode: Gas temp 300°C, flow 5 l/min, Nebulizer 30 psi, sheath gas temp 350°C, flow 12 l/min). The scan source parameters in pos and neg mode were (VCap 3500, Nozzle Voltage 500 V, Fragmentor 360, Skimmer 1 and OctopoleRFPeak 750). The data were exploratively annotated using MS-DIAL and its lipidomics database. For data integration and relative quantitation we used Lipid Data Analyzer 2.8.3_2 [32].

TCGA survival analysis

TCGA clinical data was downloaded in R Studio with the *RTCGA* and *RTCGA.clinical* packages. The clinical data was analyzed using the *survival* and *survminer* packages and plotted using the *ggsurvplot* package in ggplot2.

Parallel reaction monitoring mass spectrometry (PRM-MS)

Relative peptide amount determination. Before NanoLC-MS/MS analysis, final peptide amounts were determined by separating an aliquot of each sample on an LC-UV system equipped with a monolith column (Thermo scientific technical note 72602) and normalizing it to the peak area of 100 ng of Pierce HeLa protein digest standard (PN 88329; ThermoFisher Scientific).

NanoLC-MS/MS analysis. The nano HPLC system used was a Vanquish Neo UHPLC-System coupled to an Orbitrap Exploris 480 mass spectrometer, equipped with an Easy spray Source TNG (Thermo Fisher Scientific). Peptides were loaded onto a trap column (PepMap C18, 5 mm × 300 µm ID, 5 µm particles, 100 Å pore size, Thermo Fisher Scientific) by using 0.1% TFA. The trap column was switched in line with the analytical column (Double nanoViper™ PepMap C18, 500 mm × 75 µm ID, 2 µm, 100 Å, Thermo Fisher Scientific). The analytical column was connected to PepSep sprayer 1 (Bruker) equipped with a 10 µm ID fused silica electrospray emitter with an integrated liquid junction (Bruker, PN 1893527). Electrospray voltage was set to 2.3 kV. The analytical column flow was run at 230 nL/min, at 60 min binary gradient, with two mobile phases: water with 0.1% formic acid (A) and water with 80% acetonitrile and 0.08% formic acid (B). A and B were applied in linear gradients as follows (only B percentages reported): starting from 2% B: 2%-10% B in 10 min, 10%-24% B in 35 min, 24%-35% B in 15 min, 35%-95% B in 1 min, 95% for 5 min, and finally the column was equilibrated in 2% B for 3 analytical column volumes at 30°C (all % values are v/v; Water and ACN solvents were purchased from Thermo Scientific Price at LC-MS grade).

The Orbitrap Exploris 480 mass spectrometer was operated by a mixed MS method which consisted of one full scan (m/z range 380-1500; 15000 resolution; target value 100%) followed by the PRM of targeted peptides from an inclusion list (isolation window 0.8 m/z; normalized collision energy (NCE) 34; 30000 resolution, AGC target 200%). Spectra of unique peptides of the proteins of interest was recorded. Per protein at least 2 unique peptides were measured. The maximum injection time was set to 125 ms. Each precursor was measured in a 5 min time window. Peptides included in the PRM method are listed in **Supplementary Table 6.** A scheduled PRM method (sPRM) development, data processing and manual evaluation of results was performed in Skyline [33] (64-bit, v22.2.0.351). To derive the PRM score for each protein of interest (POI), the area of each identified peptide (Peptide area, PA) was summed and divided by the sum of all areas from all identified peptides from the three normalization proteins using the following formula:

$$\text{PRM score (POI)} = \frac{\sum \text{PA(POI)}}{\sum \text{PA(UMOD)} + \sum \text{PA(KLK1)} + \sum \text{PA(APOD)}}$$

UrineScore

The UrineScore was calculated in the following way: The PRM score of the three POIs was calculated for controls, nccRCC, and ccRCC patients. The 95% confidence interval (CI) of the median of the PRM scores was calculated for the three POIs for the controls. (i) Haptoglobin 95% CI of median = 0.003772 – 0.03508, (ii) SAA1 95% CI of median = 0 – 5.35E-9, (iii) LCN15 95% CI of median = 0 – 2.02E-5. Then, for each control and patient sample, the PRM scores for each POI was compared to the corresponding proteins 95% CI of the controls. If the PRM score was higher than the upper limit of the controls 95% CI, a value of 1 was attributed. This calculation was done for all three proteins and the values added together, meaning that each control and patient sample will receive an UrineScore which is an integer between 0-3.

Statistical analysis

All statistical analysis were performed in Prism 10 (GraphPad) or in RStudio (Posit). When comparing large omics datasets (proteomics, lipidomics or metabolomics) p values were calculated with limma-moderated Benjamini–Hochberg-corrected two-sided t-test after data processing for proteomics and metabolomics, and Benjamini-Hochberg adjusted. For comparisons of individual markers between groups, the distribution of the data was initially

determined by Shapiro-Wilk normality test. All individual marker comparisons shown in the paper did not pass the Shapiro-Wilk normality, and subsequent analysis was either performed using Mann-Whitney U-tests (two groups) or Kruskal-Wallis test with Kruskal-Wallis test with Dunn's multiple comparisons test (three groups). The performance of receiver operating characteristic curve (ROC) was assessed with the Area under ROC method (AUC) and a p value calculated by testing the null hypothesis that the AUC is equal to 0.5. Gene Ontology analysis was performed using online portals, Enrichr for proteomics data and MetaboAnalyst for metabolomics data.

Author contribution

G.J. and J.M.P conceived the study. G.J. performed and designed experiments with input and help from all co-authors as follows: M.H. with mass spectrometry sample preparation and overall data analysis; T.O. with mass spectrometry; U.L, Z.K, B.E. and M.S. with ethical approval and patient sample collection; S.M. with data analysis and interpretation; M.N. with bioinformatics; G.G. and T.K. with metabolomics; G.K. and K.S. with PRM-MS; G.N.R and T.Z. with lipidomics. G.J. and J.M.P wrote the paper with input from all authors.

Acknowledgements

Metabolomics was performed at the VBCF Metabolomics Facility which is funded by the City of Vienna through the Vienna Business Agency. Proteomics analyses were performed by the Proteomics Facility at IMP/IMBA/GMI using the VBCF instrument pool. G.J. is supported by a DOC fellowship from the Austrian Academy of Sciences. S.M. received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 841319 and the ESPRIT-Programme of the Austrian Science Fund (FWF, Project number: ESP 166). J.M.P. received funding from the Medical University of Vienna, the Austrian Academy of Sciences, the T. von Zastrow foundation, the Fundacio La Marato de TV3 (202125-31), and the Canada 150 Research Chairs Program F18-01336. We also gratefully acknowledge funding by the German Federal Ministry of Education and Research (BMBF) under the project "Microbial Stargazing - Erforschung von Resilienzmechanismen von Mikroben und Menschen" (Ref. 01KX2324).

Data availability

All analyzed omics data (proteomics from urine supernatants, proteomics from urine sediments, lipidomics and metabolomics) are available as supplementary tables with the

publication. Raw mass spectrometry proteomics data for urine supernatants, urine sediments and PRM-MS will be deposited into the PRIDE database upon favorable peer-review and made available at the date of publication.

Conflicts of interest

IMBA has filed a European patent application based on the results presented herein. G.J., T.O. and J.M.P. are inventors on this patent.

References

1. Hsieh, J.J., et al., *Renal cell carcinoma.* Nature reviews Disease primers, 2017. **3**(1): p. 1-19.
2. Muglia, V.F. and A. Prando, *Renal cell carcinoma: histological classification and correlation with imaging findings.* Radiologia brasileira, 2015. **48**: p. 166-174.
3. Giaccia, A., B.G. Siim, and R.S. Johnson, *HIF-1 as a target for drug development.* Nature reviews Drug discovery, 2003. **2**(10): p. 803-811.
4. Lonser, R.R., et al., *von Hippel-Lindau disease.* The Lancet, 2003. **361**(9374): p. 2059-2067.
5. Sato, Y., et al., *Integrated molecular analysis of clear-cell renal cell carcinoma.* Nature genetics, 2013. **45**(8): p. 860-867.
6. Hakimi, A.A., et al., *Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma.* European urology, 2013. **63**(5): p. 848-854.
7. Albiges, L., et al., *Pembrolizumab plus lenvatinib as first-line therapy for advanced non-clear-cell renal cell carcinoma (KEYNOTE-B61): A single-arm, multicentre, phase 2 trial.* The Lancet Oncology, 2023. **24**(8): p. 881-891.
8. Powles, T., et al., *Pembrolizumab versus placebo as post-nephrectomy adjuvant therapy for clear cell renal cell carcinoma (KEYNOTE-564): 30-month follow-up analysis of a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial.* The Lancet Oncology, 2022. **23**(9): p. 1133-1144.
9. Vano, Y.-A., et al., *Nivolumab, nivolumab–ipilimumab, and VEGFR-tyrosine kinase inhibitors as first-line treatment for metastatic clear-cell renal cell carcinoma (BIONIKK): A biomarker-driven, open-label, non-comparative, randomised, phase 2 trial.* The Lancet Oncology, 2022. **23**(5): p. 612-624.

10. Gray, R.E. and G.T. Harris, *Renal cell carcinoma: diagnosis and management.* American family physician, 2019. **99**(3): p. 179-184.
11. Burg, M., et al., *Organic solutes in fluid absorption by renal proximal convoluted tubules.* American Journal of Physiology-Legacy Content, 1976. **231**(2): p. 627-637.
12. Lucarelli, G., et al., *Metabolomic insights into pathophysiological mechanisms and biomarker discovery in clear cell renal cell carcinoma.* Expert review of molecular diagnostics, 2019. **19**(5): p. 397-407.
13. Zhang, Y., et al., *Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response.* Proceedings of the National Academy of Sciences, 2021. **118**(24): p. e2103240118.
14. Li, M., et al., *Liquid biopsy at the frontier in renal cell carcinoma: recent analysis of techniques and clinical application.* Molecular Cancer, 2023. **22**(1): p. 37.
15. Farber, N.J., et al., *Renal cell carcinoma: the search for a reliable biomarker.* Translational cancer research, 2017. **6**(3): p. 620.
16. Rustam, Y.H. and G.E. Reid, *Analytical challenges and recent advances in mass spectrometry based lipidomics.* Analytical chemistry, 2018. **90**(1): p. 374-397.
17. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics.* Molecular & cellular proteomics, 2012. **11**(11): p. 1475-1488.
18. Bukavina, L., et al., *Epidemiology of Renal Cell Carcinoma: 2022 Update.* Eur Urol, 2022. **82**(5): p. 529-542.
19. Diana, P., et al., *Screening programs for renal cell carcinoma: a systematic review by the EAU young academic urologists renal cancer working group.* World journal of urology, 2023. **41**(4): p. 929-940.
20. Usher-Smith, J.A., et al., *The Yorkshire Kidney Screening Trial (YKST): protocol for a feasibility study of adding non-contrast abdominal CT scanning to screen for kidney cancer and other abdominal pathology within a trial of community-based CT screening for lung cancer.* BMJ open, 2022. **12**(9): p. e063018.
21. Du, W., et al., *HIF drives lipid deposition and cancer in ccRCC via repression of fatty acid metabolism.* Nature communications, 2017. **8**(1): p. 1769.
22. Marques, L.R., et al., *Reverse cholesterol transport: molecular mechanisms and the non-medical approach to enhance HDL cholesterol.* Frontiers in Physiology, 2018. **9**: p. 526.

23. Sandim, V., et al. *Proteomic analysis reveals differentially secreted proteins in the urine from patients with clear cell renal cell carcinoma*. in *Urologic Oncology: Seminars and Original Investigations*. 2016. Elsevier.
24. Santorelli, L., et al., *In-depth mapping of the urinary N-glycoproteome: distinct signatures of ccRCC-related progression.* Cancers, 2020. **12**(1): p. 239.
25. Chinello, C., et al., *Proteomics of liquid biopsies: Depicting RCC infiltration into the renal vein by MS analysis of urine and plasma.* Journal of proteomics, 2019. **191**: p. 29-37.
26. Yang, Y., et al., *Excavation of diagnostic biomarkers and construction of prognostic model for clear cell renal cell carcinoma based on urine proteomics.* Frontiers in Oncology, 2023. **13**: p. 1170567.
27. Di Meo, A., et al., *Searching for prognostic biomarkers for small renal masses in the urinary proteome.* International journal of cancer, 2020. **146**(8): p. 2315-2325.
28. Wessel, D. and U. Flügge, *A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids.* Analytical biochemistry, 1984. **138**(1): p. 141-143.
29. Bruderer, R., et al., *Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues*[S].* Molecular & Cellular Proteomics, 2015. **14**(5): p. 1400-1410.
30. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Statistical applications in genetics and molecular biology, 2004. **3**(1).
31. Matyash, V., et al., *Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics.* Journal of lipid research, 2008. **49**(5): p. 1137-1146.
32. Hartler, J., et al., *Deciphering lipid structures based on platform-independent decision rules.* Nature methods, 2017. **14**(12): p. 1171-1174.
33. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-968.

Supplementary tables

Supplementary Table 1. Patient information. Table containing information regarding patient age, gender, clinical parameters (e.g. C-reactive protein levels in serum), type of tumor and stage of tumor at the timepoint of urine collection.

Supplementary Table 2. Supernatant proteomics. Table containing raw expression protein data (normalized area) for each detected protein in the urine supernatant for each sample.

Supplementary Table 3. Sediment proteomics. Table containing raw expression protein data (normalized area) for each detected protein in the urine sediment for each sample, as well as group averages and statistics.

Supplementary Table 4. Metabolomics. Table containing raw expression metabolite data (normalized area) for each detected metabolite in the urine for each sample.

Supplementary Table 5. Lipidomics. Table containing raw expression lipid data (normalized area) for each detected lipid class and species in the urine for each sample.

Supplementary Table 6. PRM peptides. Table containing peptide sequences for the diagnostic proteins (SAA1, HP and LCN15) and the normalization proteins (UMOD, KLK1 and APOD) used in the PRM-MS approach.

Figures and figure legends

Figure 1

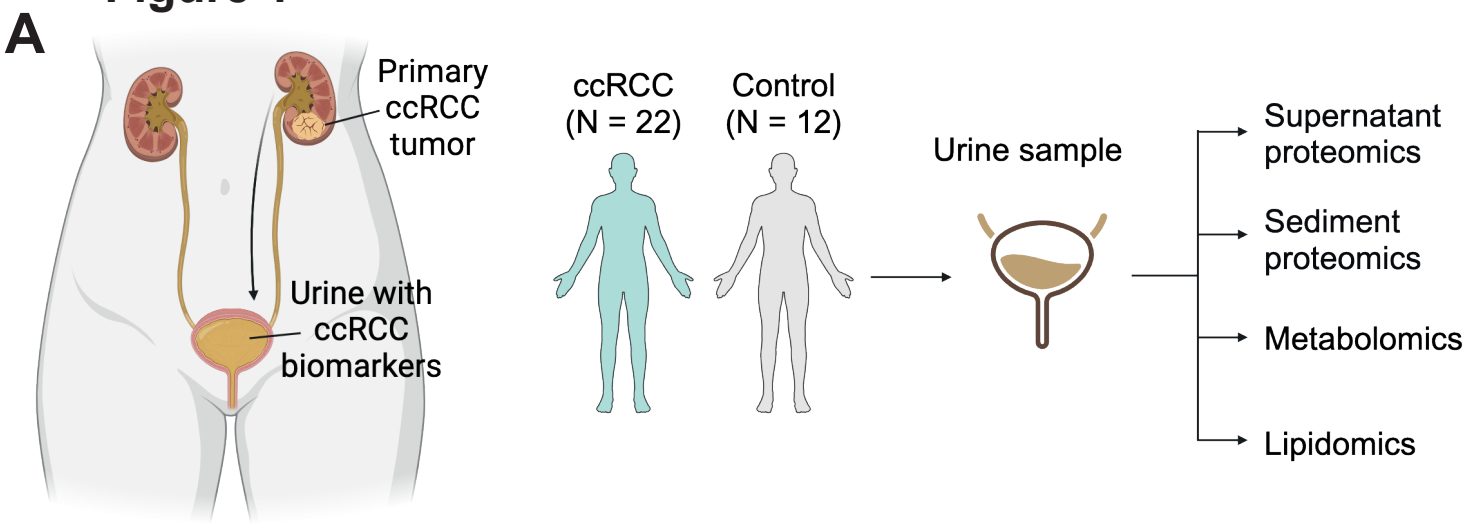


Figure 1. Multi-omics analysis of ccRCC patient urine indicates metabolic dysregulation.

A. Schematic of detection of urine analytes in ccRCC patients passing through the urogenital tract and accumulating in the urine bladder before discharge in a ccRCC patient cohort. Schematic created with BioRender.com.

B. Gene ontology (GO) term analysis using Enrichr on all upregulated proteins in ccRCC urine supernatant compared to Control urine supernatant. FDR = False discovery rate.

C. Gene ontology (GO) term analysis using Enrichr on all upregulated proteins in ccRCC urine sediment compared to Control urine sediment. FDR = False discovery rate.

D. Gene ontology (GO) term analysis using Metaboanalyst on all downregulated metabolites in ccRCC urine supernatant compared to Control urine supernatant. The downregulated metabolites were matched against two different databases: KEGG and SMPDB. FDR = False discovery rate.

E. Heatmap showing fold change between the detected lipid families in ccRCC patient urine and control urine. CoQ10 = Coenzyme Q10, PC = Phosphatidylcholines, PE = Phosphatidylethanolamine, PS = Phosphatidylserine, Chol = Cholesterol, CAR = Acylcarnitines, PE O = Ether-linked phosphatidylethanolamine, TG = Triglyceride, SM = Sphingomyelin.

F. Boxplot of individual values for the three significantly higher lipid families from **E**. CoQ10 = Coenzyme Q10, PC = Phosphatidylcholines, PE = Phosphatidylethanolamine. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

G. Schematic of proposed altered metabolic network in the urogenital tract of ccRCC patients. Green represents upregulated pathways and red represents downregulated pathways. HDL = High-density lipoprotein, CoA = Coenzyme A, α-KG = alpha-ketoglutarate, CoQ10 = Coenzyme Q10. Schematic created with BioRender.com.

Figure 2

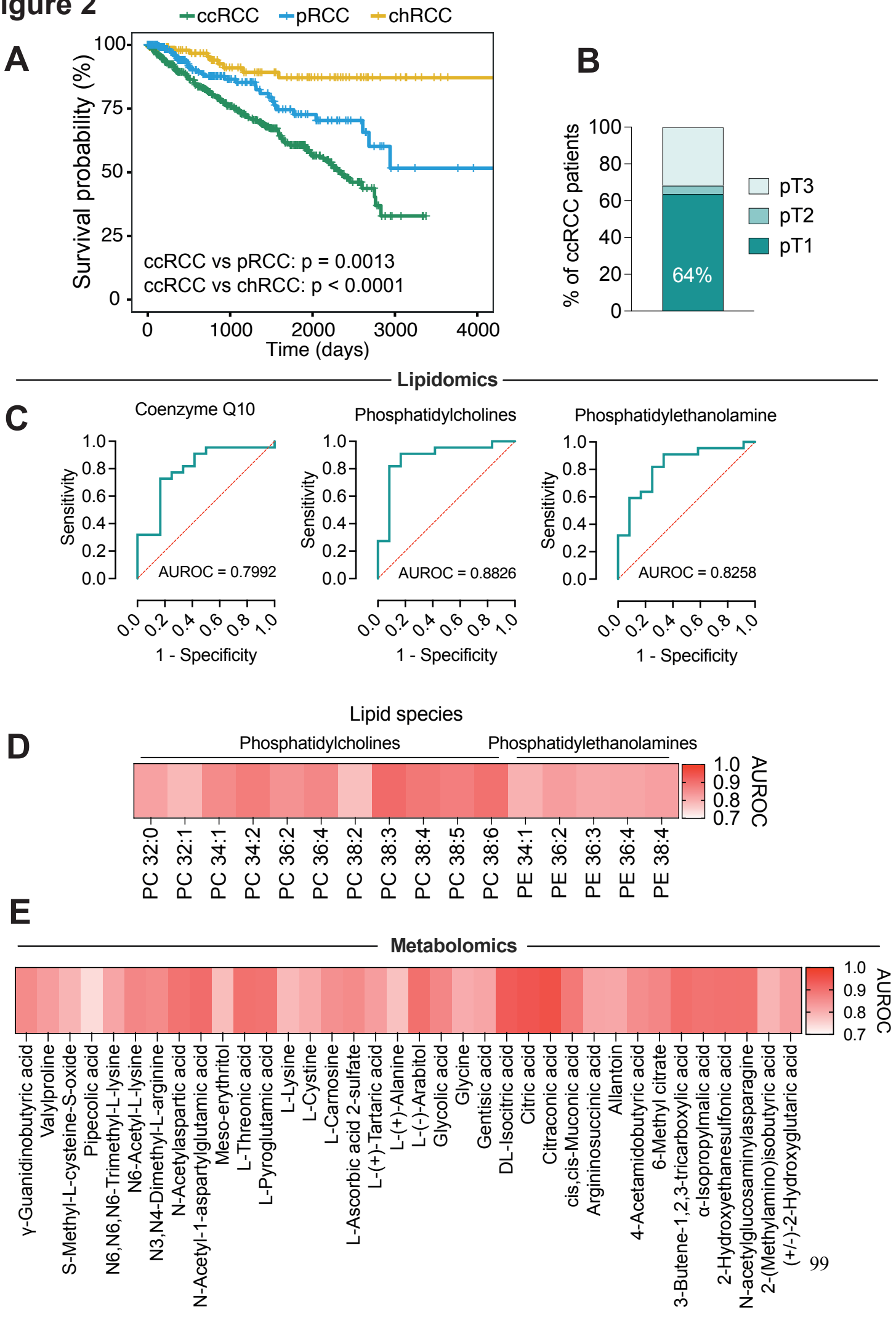


Figure 2. Lipidomics and metabolomics reveal biomarkers that distinguish between controls and ccRCC patients.

A. Survival curves ccRCC, pRCC and chRCC patients present in the TCGA database. TCGA = The Cancer Genome Atlas, ccRCC = clear cell renal cell carcinomas, pRCC = papillary renal cell carcinomas, chRCC = chromophobe renal cell carcinoma. Statistical significance assessed through Log-rank tests.

B. Distribution of tumor stages at the timepoint of presentation in the clinic for ccRCC patients. Abbreviations: pT1 = Tumor stage 1. The tumor is a maximum of 7 cm across. pT2 = Tumor stage 2. The tumor is larger than 7 cm across. pT3 = Tumor stage 3. The tumor has grown into a major renal vein (e.g. vena cava or renal vein) or into neighboring tissues but has not spread past Gerota's fascia or into the adrenal gland.

C. Assessment of the ability of the three significantly upregulated lipid families (CoQ10, PC and PE, respectively) in distinguishing between ccRCC and control cases using AUROC analysis. CoQ10 = Coenzyme Q10, PC = Phosphatidylcholines, PE = Phosphatidylethanolamine, AUROC = Area under receiver operating characteristic curve.

D. Heatmap showing AUROC values for individual lipid species from PC and PE that were upregulated in ccRCC compared to controls. PC = Phosphatidylcholines, PE = Phosphatidylethanolamine, AUROC = Area under receiver operating characteristic curve.

E. Heatmap showing AUROC values for individual metabolites that were downregulated in ccRCC compared to controls. AUROC = Area under receiver operating characteristic curve.

Figure 3

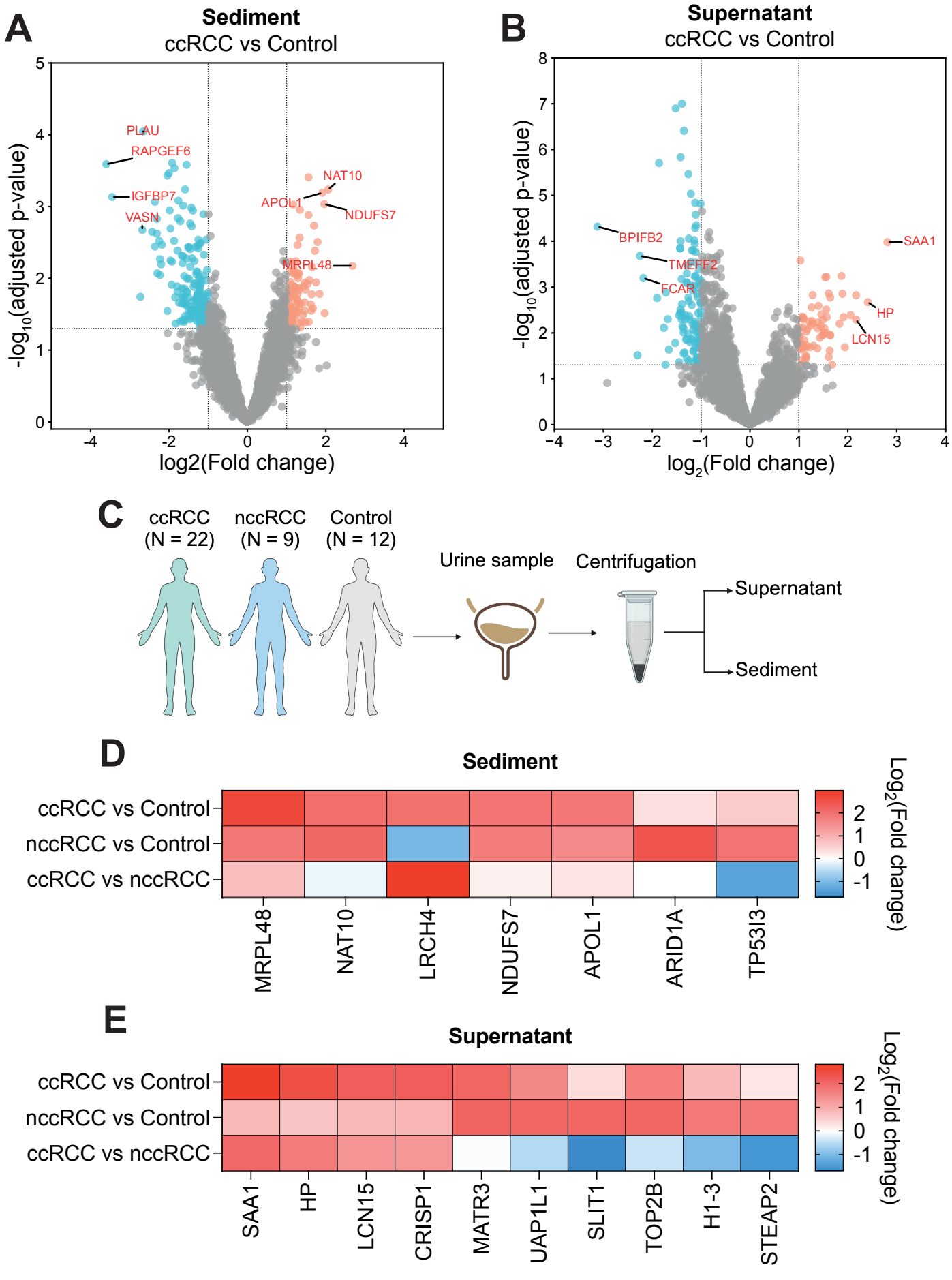


Figure 3. Identification of protein biomarkers for ccRCC diagnosis.

A. Volcano plot showing upregulated (red) and downregulated (blue) proteins in ccRCC patients compared to controls in urine sediment. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test. FC = fold change.

B. Volcano plot showing upregulated (red) and downregulated (blue) proteins in ccRCC patients compared to controls in urine supernatant. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test. FC = fold change.

C. Schematic of sample acquisition from ccRCC patients, nccRCC patients and controls for urine supernatant and sediment proteomics.

D. Heatmap of the highest upregulated proteins in ccRCC vs control urine (top row) and nccRCC vs control urine (middle row) in urine sediments. A comparison between ccRCC and nccRCC is also shown in the bottom row.

E. Heatmap of the highest upregulated proteins in ccRCC vs control urine (top row) and nccRCC vs control urine (middle row) in urine supernatans. A comparison between ccRCC and nccRCC is also shown in the bottom row.

Figure 4

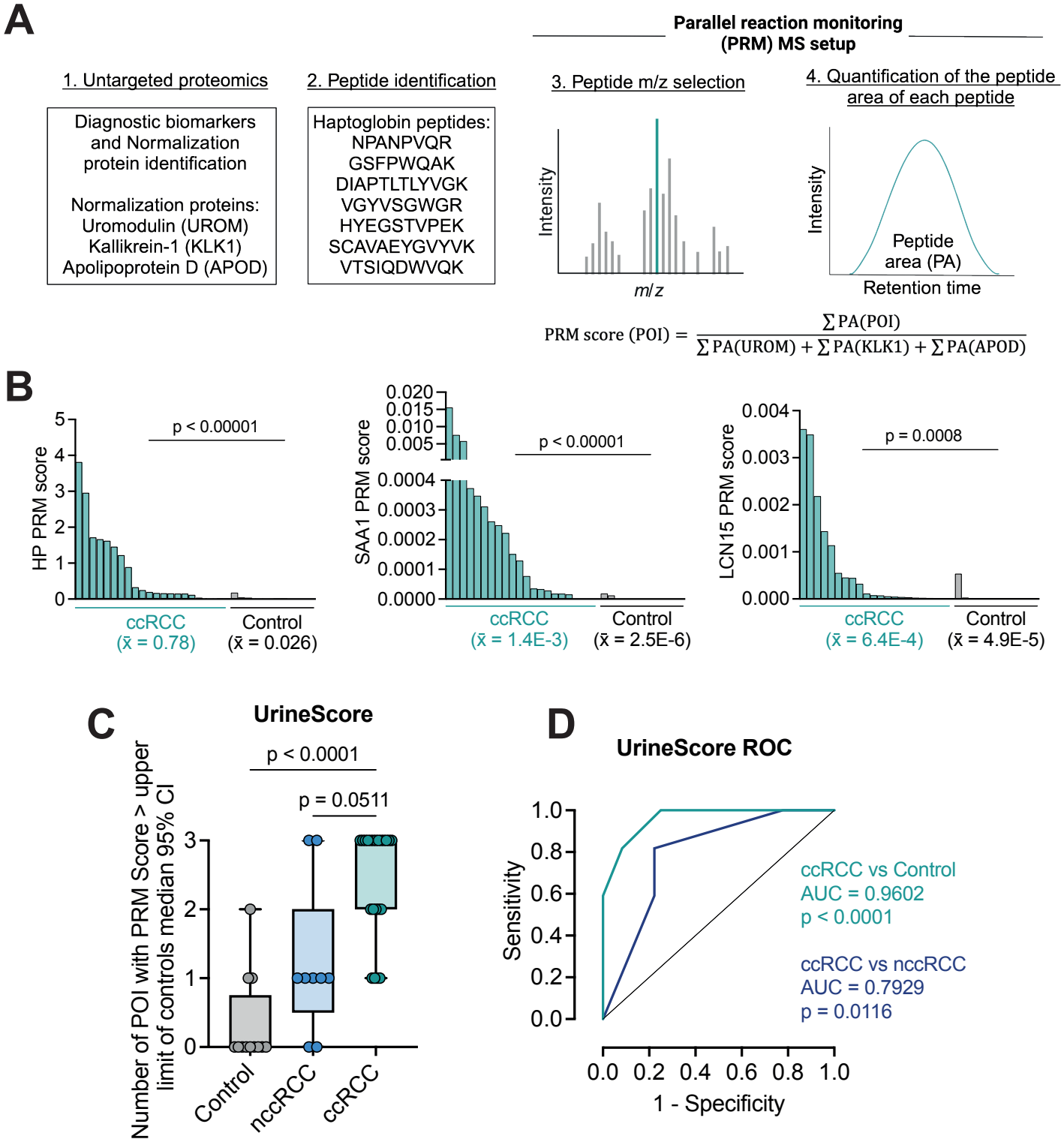


Figure 4. Parallel reaction monitoring mass spectrometry accurately diagnoses ccRCC.

A. Schematic of the PRM-MS method used to quantify the SAA1, HP and LCN15 levels in urine samples. Levels are determined based on normalization to three normalization proteins (UMOD, KLK1 and APOD). The peptide area for each detected peptide from a specified protein is quantified, detected peptides from Haptoglobin are shown as an example. Abbreviations: PA = Peptide area, POI = Protein of interest (SAA1, HP or LCN15). Schematic created using BioRender.com.

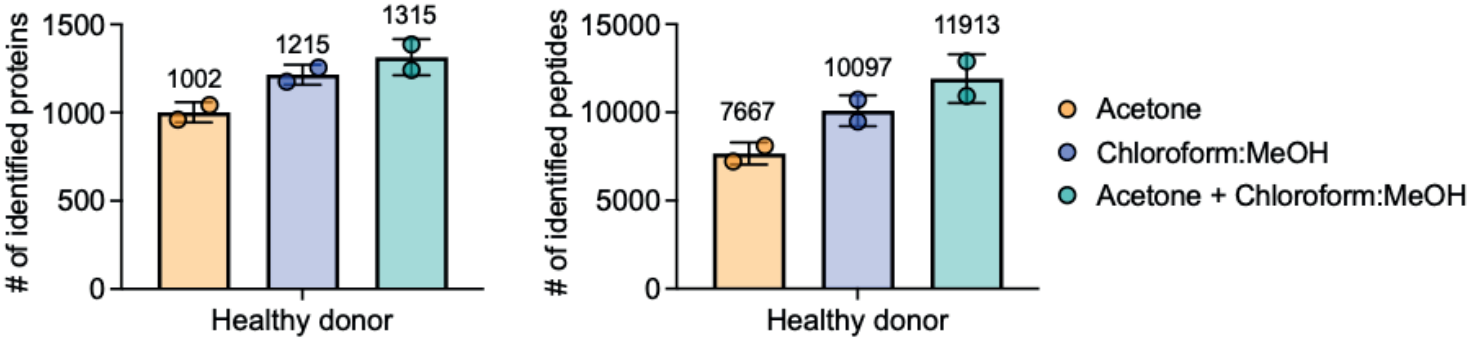
B. Waterfall plot of PRM score for Haptoglobin, SAA1 and LCN15, respectively, compared between ccRCC and Control cohorts. Statistical significance assessed through Mann-Whitney U-test. $\bar{x}$ = sample mean.

C. Box plot of UrineScore for controls, nccRCC and ccRCC patients. Per protein of interest (POI) per patient or control, a value of 1 is derived if the PRM score is higher than the upper limit of the 95% CI for the control. Statistical significance assessed through Kruskal-Wallis test with Dunn's multiple comparisons test.

D. Receiver operating characteristic curve to assess the performance of the UrineScore in differentiating between ccRCC and control samples, and between ccRCC and nccRCC samples. Performance quantified through Area under the receiver operating characteristic curve (AUROC) analysis.

Supplementary Figure 1

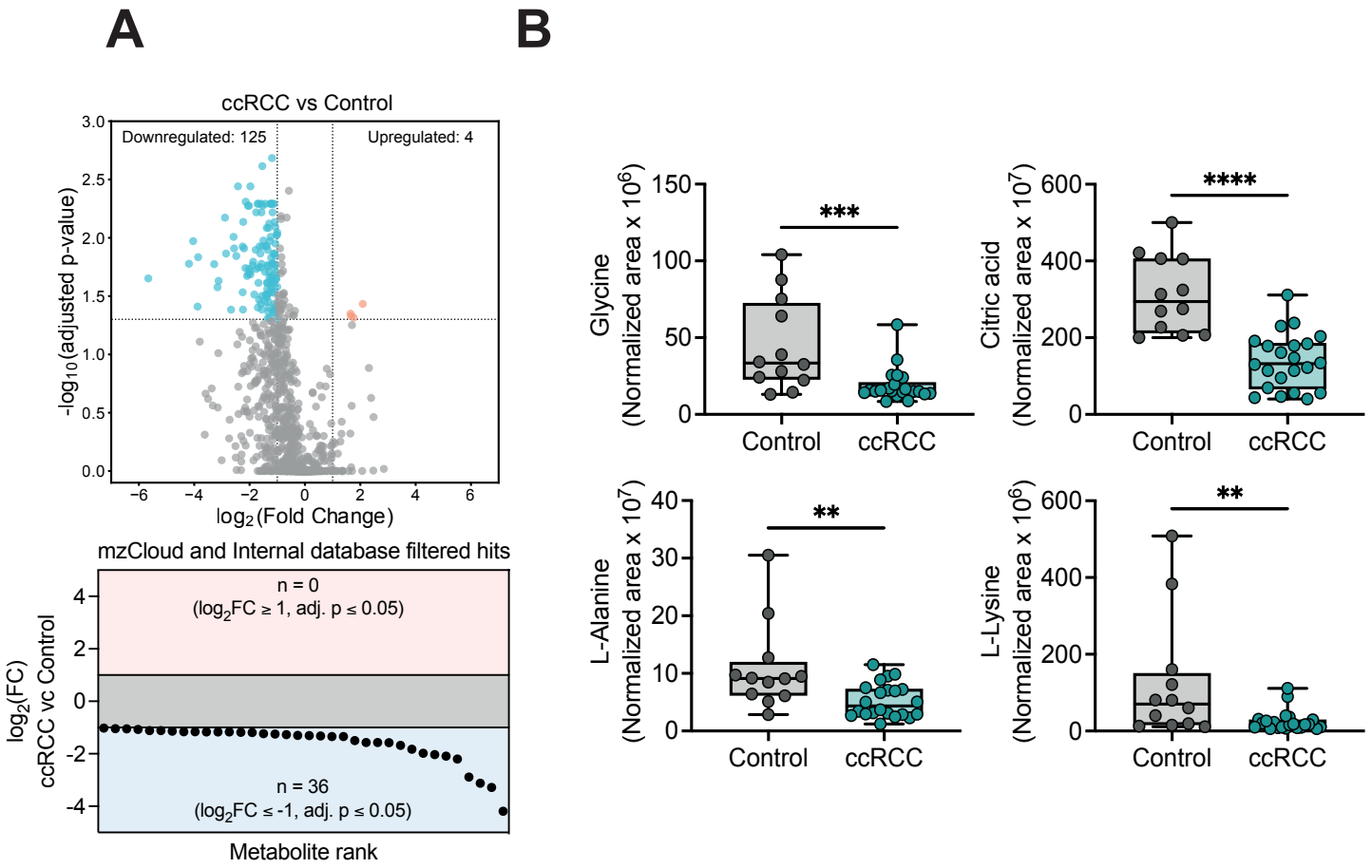
A



Supplementary Figure 1. Double protein precipitation increases unique peptide and protein yields from urine supernatant.

Number of identified unique proteins (left) and unique peptides (right) in a healthy control urine supernatant sample using three different protein precipitation methods. (i) Precipitation only with acetone, (ii) precipitation only with chloroform and methanol, and (iii) double protein precipitation with acetone followed by chloroform and methanol. Two technical replicates per sample is shown. MeOH = Methanol.

Supplementary Figure 2



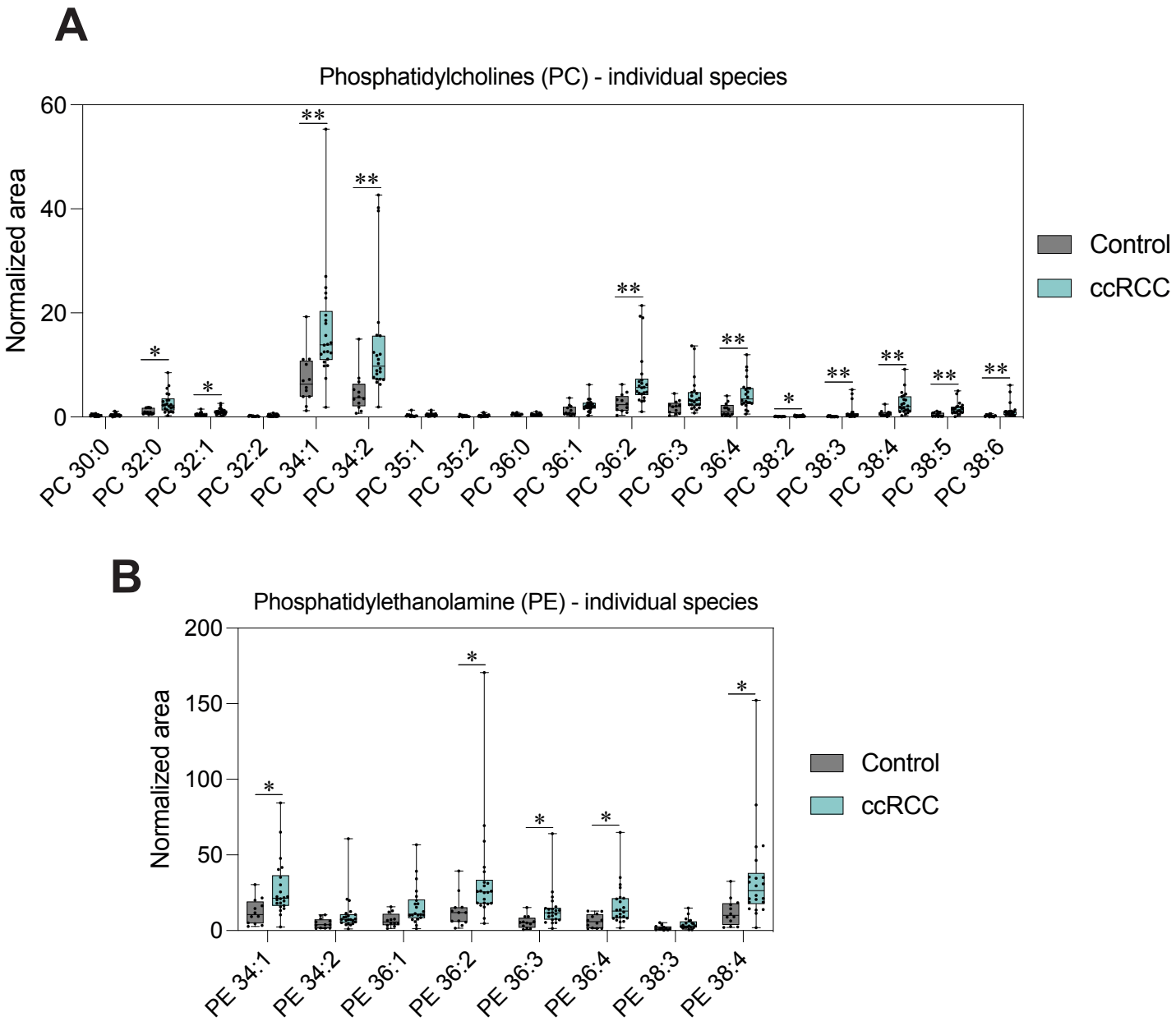
Supplementary Figure 2. Metabolites in the urine of ccRCC patients.

A. Volcano plot showing upregulated (red) and downregulated (blue) metabolites in ccRCC patients compared to controls. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test. FC = fold change.

B. Waterfall plot of significantly metabolites remaining when comparing ccRCC and control urine after filtering all hits from **A** through the mzCloud and an internal database. 36 significantly downregulated metabolites remain.

C. Boxplots of 4 representative metabolites expression values (Glycine, Citric acid, L-Alanine and L-Lysine) as determined by metabolomics. ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Supplementary Figure 3

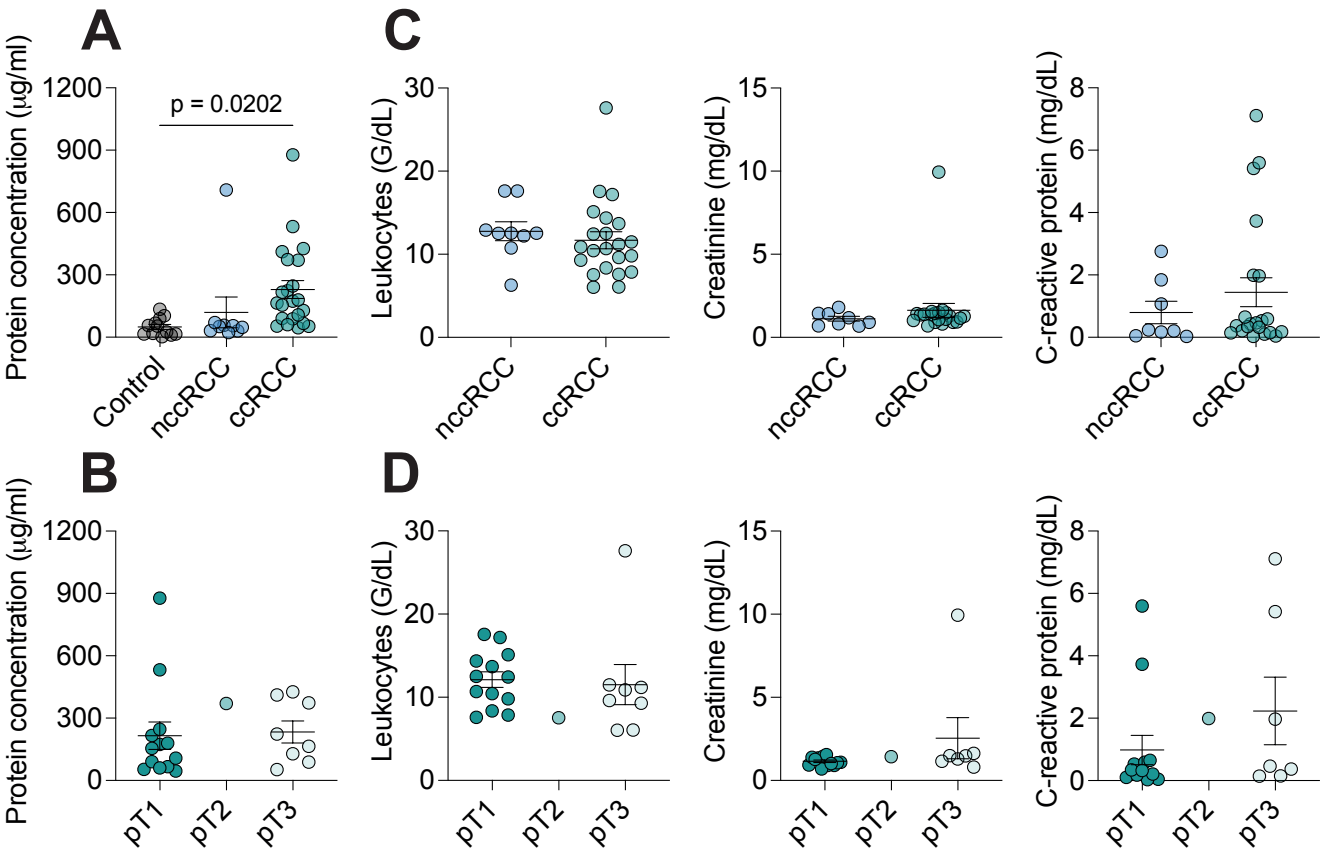


Supplementary Figure 3. PC and PE lipid species are enriched in ccRCC urine.

A. Boxplots of individual lipid species from the PC family. PC = Phosphatidylcholines. * = $p < 0.05$, ** = $p < 0.01$.

B. Boxplots of individual lipid species from the PE family. PE = Phosphatidylethanolamine. * = $p < 0.05$.

Supplementary Figure 4



Supplementary Figure 4. Clinical cohort overview data.

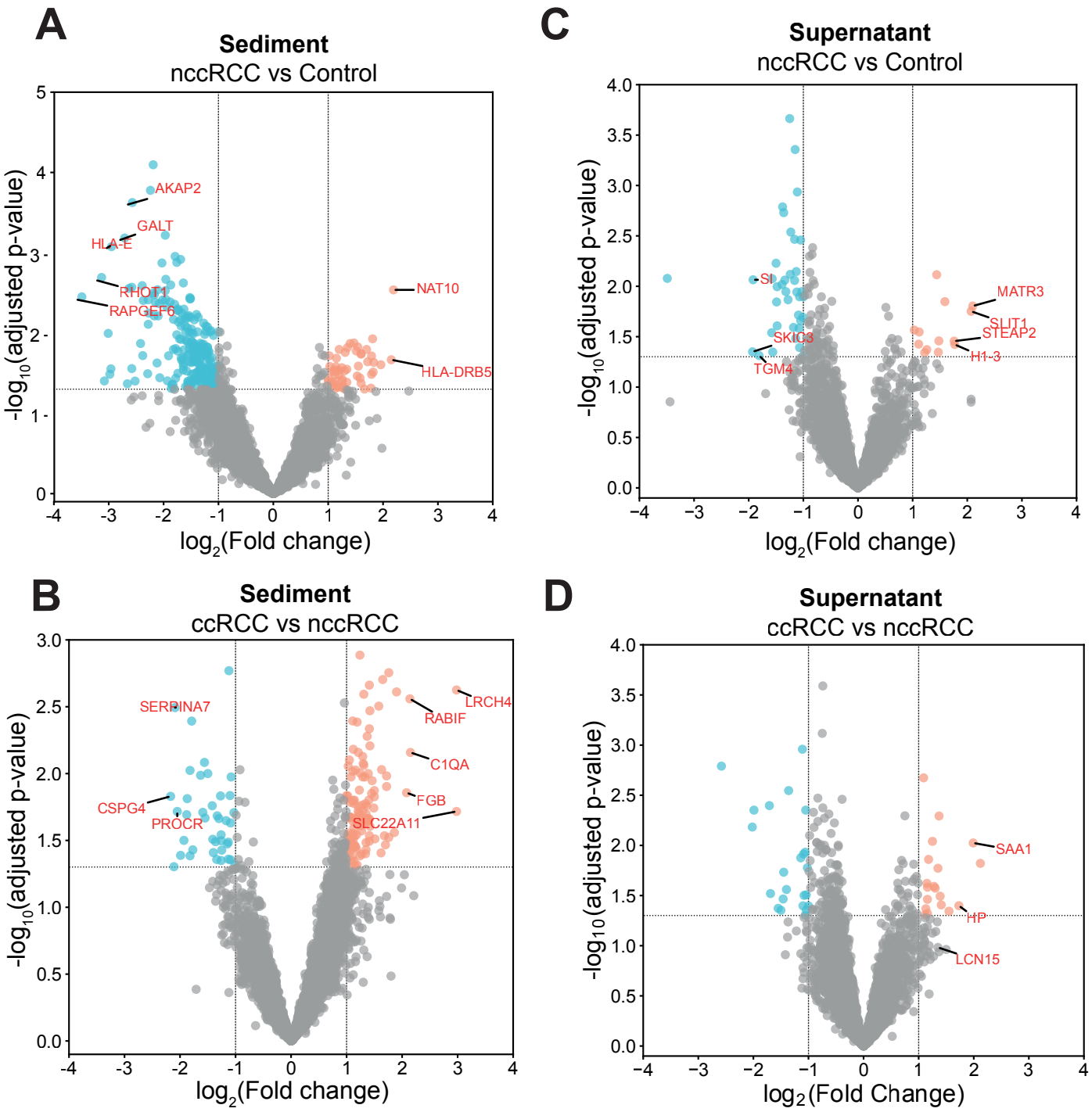
A. Protein concentration in urine supernatant from control, nccRCC patients and ccRCC patients determined via Bradford protein assay. * = $p < 0.005$

B. Protein concentration in urine supernatant from ccRCC patients divided into pT1, pT2 and pT3 stages determined via Bradford protein assay. Abbreviations: pT1 = Tumor stage 1. The tumor is a maximum of 7 cm across. pT2 = Tumor stage 2. The tumor is larger than 7 cm across. pT3 = Tumor stage 3. The tumor has grown into a major renal vein (e.g. vena cava or renal vein) or into neighboring tissues but has not spread past Gerota's fascia or into the adrenal gland.

C. Leukocyte, Creatinine and C-reactive protein levels in serum, respectively, for nccRCC patients and ccRCC patients

D. Leukocyte, Creatinine and C-reactive protein levels in serum, respectively, ccRCC patients divided into pT1, pT2 and pT3 stages. Abbreviations: pT1 = Tumor stage 1. The tumor is a maximum of 7 cm across. pT2 = Tumor stage 2. The tumor is larger than 7 cm across. pT3 = Tumor stage 3. The tumor has grown into a major renal vein (e.g. vena cava or renal vein) or into neighboring tissues but has not spread past Gerota's fascia or into the adrenal gland.

Supplementary Figure 5



Supplementary Figure 5. Distinct urinary protein landscapes in ccRCC and nccRCC patients.

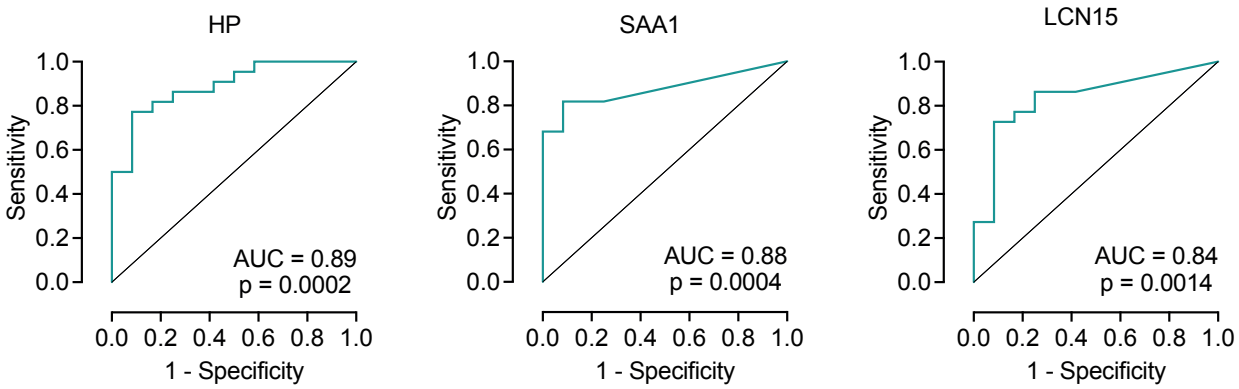
A. Volcano plot showing upregulated (red) and downregulated (blue) proteins in nccRCC patients compared to Controls in urine sediment. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test.

B. Volcano plot showing upregulated (red) and downregulated (blue) proteins in nccRCC patients compared to ccRCC patients in urine sediment. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test.

C. Volcano plot showing upregulated (red) and downregulated (blue) proteins in nccRCC patients compared to controls in urine supernatant. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test. FC = fold change.

E. Volcano plot showing upregulated (red) and downregulated (blue) proteins in ccRCC patients compared to nccRCC patients in urine supernatant. Adjusted p values calculated via limma-moderated Benjamini–Hochberg-corrected two-sided t-test. FC = fold change.

Supplementary Figure 6



Supplementary Figure 6. PRM-MS enables detection of SAA1, HP and LCN15 in the urine of ccRCC patients.

Receiver operating characteristic curve to assess the performance of the PRM score of Haptoglobin, SAA1 and LCN15, respectively, in differentiating between control and ccRCC samples. Performance quantified through Area under the receiver operating characteristic curve (AUC) analysis.

4. Discussion

Clear cell renal cell carcinoma (ccRCC) represents a significant clinical challenge due to its rising incidence and mortality rates [137]. With the rapid development of new or advanced therapeutics for cancer, the treatment landscape for ccRCC has improved in the last couple of years, especially with the introduction of immunotherapies [138]. Despite these recent developments, ccRCC treatment still faces several issues, such as the rapid development of treatment resistance to first-line therapies and the sometimes late discovery of the primary tumor meaning that patients present in the clinic with already advanced disease [77, 82, 139]. There is therefore a clinically unmet need for better early stage diagnostic and prognostic biomarkers for informed treatment strategies at an early stage of the disease.

One promising area for discovery of screenable markers are liquid biopsies which have rapidly gained traction in the cancer field in the last decade. Liquid biopsies are body fluids, e.g. blood or urine, in which disease markers are probed, predominantely tumor cells or cancer cell DNA [140, 141]. Compared to other types of cancer, the implementation of liquid biomarkers in ccRCC has been slow. Currently, few options are available and the most promising diagnostic liquid biopsy biomarker at the moment are DNA-containing extracellular vesicles, but these results are still missing essential validations [142]. In this thesis, I aimed to discover and validate both prognostic and diagnostic biomarkers for ccRCC in liquid biopsies which resulted in two publications.

In an initial *in silico* screening of secreted genes (**Fig. 7**) we discovered a gene family consisting of three paralogs (CLEC18A, CLEC18B and CLEC18C) which were all upregulated in ccRCCs compared to healthy surrounding tissue. The CLEC18 paralog family is poorly characterized and its role in kidney cancer has remained undescribed. TCGA data showed that the CLEC18 family carries prognostic value in ccRCC specifically, and no other types of kidney cancer. These results are demonstrated in **Publication 1**. In the same *in silico* screen, we also found three genes (SAA1, HP and LCN15) with potential diagnostic value in ccRCC. To investigate them further, we performed proteomics on urine supernatants and sediments from ccRCC patients, as well as metabolomics and lipidomics, to validate them as diagnostic biomarkers. We also implemented a rapid RPM-MS protocol for sensitive detection of SAA1, HP and LCN15 in urine supernatants. These results are described in **Publication 2**. The main tumor-related findings from **Publication 1** and **Publication 2** are shown in **Fig. 8** and the results of each publication are further discussed below.

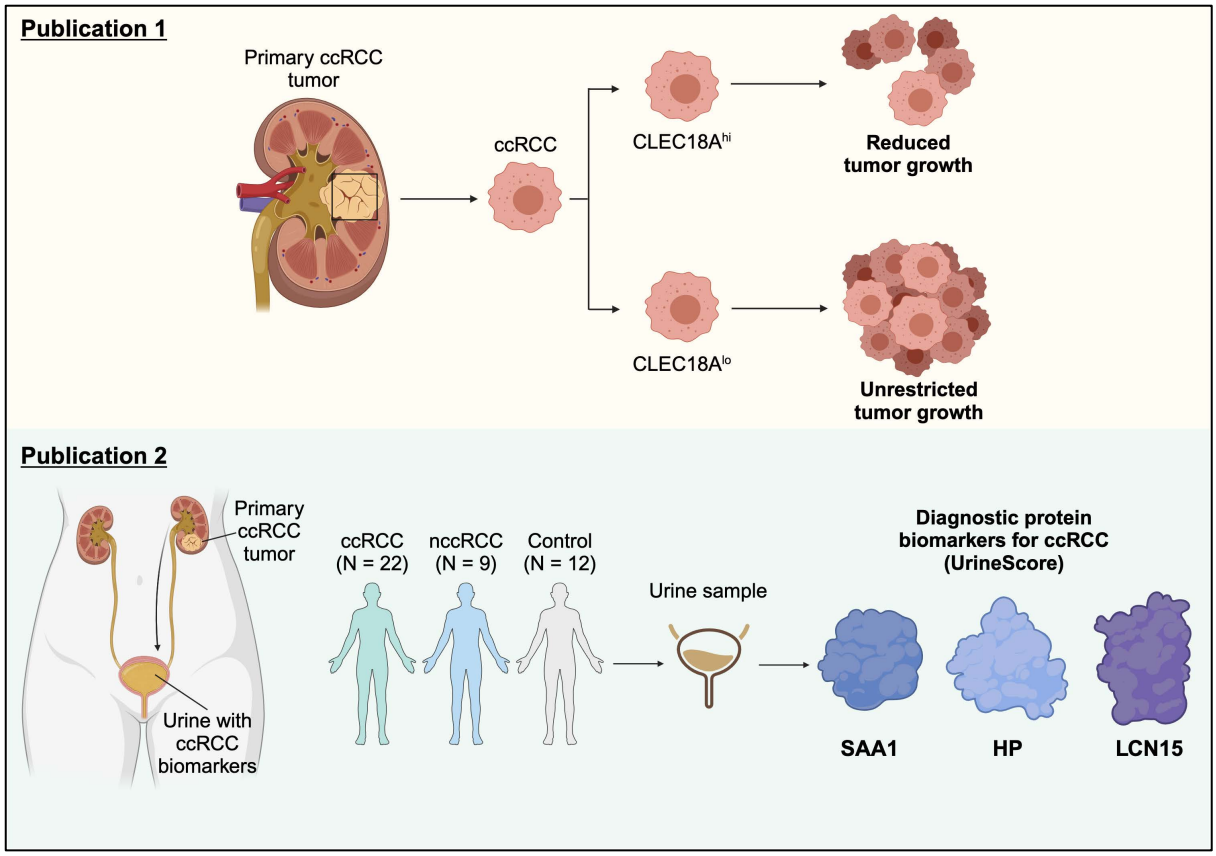


Figure 8. Summary of the main tumor results from Publication 1 and Publication 2. In **Publication 1**, upper part, it was uncovered that expression of the gene family *CLEC18A* positively correlates with tumor control in both human datasets (TCGA) and mouse models of ccRCC. In **Publication 2**, in a discovery cohort, SAA1, HP and LCN15 were shown to have diagnostic value for early detection of ccRCC. These three protein markers were combined into a cumulative UrineScore which could distinguish between ccRCC and controls, and between ccRCC and nccRCC. Schematic created with BioRender.com.

4.1 CLEC18 paralogs as prognostic biomarkers for ccRCC

Previously, the main information that has been generated about the CLEC18 family has implicated that the proteins encoded in this gene cluster have glycan binding activity, like other members of the C-type lectin family, but the ligand has remained elusive [143]. Furthermore, recent studies have shown that the gene cluster has paralogous sequence variants and polymorphic variants in humans [144]. In a disease context, correlative links have been found between CLEC18 levels and Hepatitis C [145] and Hepatitis B [146] disease severity. Lastly, a missense *cis*-acting SNP in CLEC18A (rs75776403) was implicated to lead to immune deactivation and altered kidney traits [147].

Due to the limited available information regarding the CLEC18 family in both physiological states and disease, especially with regards to malignancy, we first attempted to generate more

general information about CLEC18. Evolutionarily, we initially found that CLEC18 is highly conserved across the entire Chordata phylum dating back to lampreys with a variable number of paralogs per species. Our phylogenetic data highlighted evolutionarily recent mutational events, most likely duplications, in monkeys and humans. For example, Gorillas and higher primates have two isoforms of CLEC18 whereas humans have three. The exact reason for the high mutation rate in this locus remains unknown, but as far as we can tell the paralogs in human are redundant. In terms of expression, we found *CLEC18A* to be highly restricted to the proximal tubule in the kidney and the medial habenula in the brain, with higher expression in the proximal tubule. The medial habenula is an understudied part of the brain. It makes up one of the subunits of the habenula which is a bilateral neuronal structure connecting regions of the forebrain with regions of the midbrain. Little is known about the function of the medial habenula but it has been found to be involved in the regulation of anxiety and addiction, and addiction-related mood-regulation [148-150]; what, if any, role CLEC18 might serve in these pathways remains to be discovered.

To get some clues regarding the function of CLEC18, we attempted to resolve its ligand. Initially, phylogenetic data of the carbohydrate recognition domain (CRD) of CLEC18A showed that the amino acid sequence of the CRDs was unique amongst C-type lectins. Further, through a combination of lectin-mediated pulldowns, alphafold interaction screens and glycoarrays we managed to elucidate its ligand to be sulfated glycosaminoglycans (GAGs), such as heparin, heparan sulfate and chondroitin sulfate, anchored to proteoglycans, such as Syndecan 1, Syndecan 2, Syndecan 4, Versican, Glypican 1 or Biglycan. To further validate the GAG binding properites of CLEC18A, enzymatic degradation of GAGs can be performed. We know which cell lines our CLEC18A-Fc construct binds to, specific GAGs can be cleaved from their surfaces to confirm binding depletion. To determine the steady state kidney and brain function of CLEC18, and the role of its GAG binding activity, a knockout mouse would prove an invaluable tool. Another complementary tool would be *Clec18a*$^{-/-}$ stem cells which can be differentiated into kidney organoids. For all genetic manipulation approaches, murine systems would be the easiest starting point due to the differences in paralog numbers (mouse has one paralog, humans have three) due to the presumed redundancy, meaning that the simultaneous manipulation of multiple targets would be required in human systems.

In a cancer state, by analyzing *CLEC18* expression data across every cancer type present in TCGA we found that *CLEC18* is only associated with survival outcome in kidney cancer and brain cancer, the two organs where it is expressed. More specifically, in kidney cancer high *CLEC18* expression is correlated with better survival in ccRCC. Interestingly, *CLEC18*

expression levels apparently did not correlate with the survival outcome in pRCC or chRCC. Morever, whereas *CLEC18* is upregulated compared to normal tissue in ccRCC, this was not observed in pRCC or chRCC where *CLEC18* mRNA expression was found to be downregulated. These findings were also replicated in mouse models in which *Clec18a*$^{-/-}$ ccRCC cells subcutaneously injected into mice grew out and reached the experimental endpoint faster than wildtype cell counterparts, confirming the protective role of CLEC18 in the ccRCC tumors. In summary, in ccRCC *CLEC18* is generally upregulated in the tumor. However, despite the upregulation in the tumor, higher expression levels are associated with a better prognosis in terms of survival for the patients. The exact mechanisms by which CLEC18 protects against tumor progression remains elusive.

Having discovered the prognostic role of CLEC18 in ccRCC, and having investigated if the CLEC18 family can theoretically be used prognostically in ccRCC, we next measured the protein levels of CLEC18 in the urine discovery cohort used for **Publication 2**. However, in the mass spectrometry data from all urine samples in the study, i.e. 22 ccRCC, 9 non-clear cell RCC (nccRCC) and 12 controls, not a single peptide from the CLEC18 family was detected in any of the samples. According to the Expasy PeptideCutter tool [151], using Trypsin and LysC as tryptic enzymes, multiple theoretical peptides should be available. There are many possible explanations as to why none of these were detected in our urine samples. For example, the baseline expression of CLEC18 is low, and even with the upregulation in ccRCC it is not enough to surpass the limit of detection in mass spectrometry. CLEC18 contains signal peptides and is thus implied to be secreted from the proximal tubule where it is expressed. It is possible that it is exclusively secreted into circulation and not into the excreted urine fraction, or that it is secereted and immediately binds back to the secreted cell or in anchors in the extracellular matrix through its GAG ligands. Of course, general tissue damage in ccRCC could lead to the displacement of CLEC18 into the urine but then with generally low expression levels it would still fall under the detection limit. Furthermore, as a GAG binder, it might find abundant ligands in the urinary tract and ends up being retained and also possibly uptaken throughout the urogenital tract through selective uptake mechanisms. Lastly, urine is high in salts, organic compound (such as urea) and low in pH. The formation of amino acid adducts, and chemical modifications could lead to the failed discovery of the CLEC18 peptides since mass spectrometry software are often blind to unexpected modification.

Despite the failed validation of CLEC18 as assayable urine prognostic biomarkers in ccRCC, it does not exclude it as a putative prognostic marker nor decrease the value of studying it in a ccRCC context. Much remains to be understood about how CLEC18 and how its GAG ligands

operate in steady state and disease. Other liquid biopsies, such as serum, could provide measureable CLEC18 levels for patient stratification.

4.2 SAA1, HP and LCN15 as diagnostic biomarkers for ccRCC

In order to find, and validate diagnostic biomarkers from the *in silico* screen (**Fig. 7**, **Fig. 8**), we recruited a small-scale discovery patient cohort at the Medical University of Vienna consisting of 22 ccRCC patients and 12 controls. The patient cohort included mainly early stage 1 tumors facilitating the discovery of early diagnostic biomarkers of ccRCC. Urine was collected from all patients and controls, and for the patients the urine was collected immediately after the primary tumor was discovered and before any partial or radical nephrectomy was performed.

The urine samples were processed and subjected to proteomics, lipidomics and metabolomics. Proteomics was performed on both the supernatant and sediment urine fractions. Combining the data from all omics approaches initally provided us with a novel biological information regarding the metabolic state throughout the urogenital tract in ccRCC patients. Combined, the ccRCC urine multi-omics indicated increased lipid content and lipid processes, as well as increased mitochondrial respiration. The lipid phenotypes, most likely, stem directly from the tumor due to the general lipid accumulation taking place in ccRCC cells making them lipid-laden [152]. The increased mitochondrial respiration, on the other hand, is most likely coming from healthy kidney cells in the urine sediments, since ccRCC tends to be independent of oxidative phosphorylation and rely more on the Warburg effect [117, 153]. As a hypothetical scenario, tissue damage from the growing tumor would cause more healthy cells to be displaced and thereby end up in the urine.

With the observed phenotypes pertaining to shifted metabolic pathways, many molecular analytes related to the abrogated pathways also showed dysregulation and could be assayed for diagnostic purposes. Lipidomics, metabolomics and proteomics all showed similar potential for diagnosing ccRCC. However, lipidomics and metabolomics remains more challenging than proteomics with dedicated machine setups and analysis workflows. We therefore focused on proteomics, also because it is the most common clinically available workflow out of the three and is routinely performed in many clinics [154, 155]. Furthermore, protein-based readouts also open up the avenues of developing antibody-based methods in the future, such as ELISAs which are rapid and often used in an automated fashion in routine diagnostic practices [156].

To further test the validity of urinary proteomics to diagnose ccRCC, we extended our discovery cohort to also include 9 nccRCC patients since we not only want to be able to distinguish ccRCC from controls, but also other subtypes of RCC. The overlap of upregulated proteins between nccRCC and ccRCC was high in the sediment, making efficient subtype differentiation challenging, as compared to the supernatants in which ccRCC and nccRCC were more separated. Therefore, our diagnostic efforts were focused on the supernatant. SAA1, HP and LCN15 were identified as high-scoring diagnostic biomarkers and a parallel reaction monitoring mass spectrometry (PRM-MS) allowing rapid and sensitive quantification of only the three proteins of interest was implemented [157].

The PRM-MS data from the three proteins of interest was combined into a UrineScore which had a 96% performance accuracy, AUROC, in differentiating between ccRCC and controls, and 79% in differentiating between ccRCC and nccRCC. Overall, the UrineScore performs better than, or comparably to, other urinary protein-based ccRCC diagnosis methods [97, 113]. However, the study to which the UrineScore performed comparably in terms of AUROC accurately diagnoses both ccRCC and pRCC compared to controls, but cannot distinguish between them [113] showcasing the power of our UrineScore. Despite these promising early results, our nccRCC cohort was very small (n = 9, 1 patient with chRCC and 8 patients with pRCC). To assess how well the UrineScore can distinguish between subtypes of RCC higher numbers of diverse subtypes of RCC need to be evaluated. Furthermore, to truly determine the performance of the UrineScore, validations need to be performed in independent clinical cohorts from other clinical biobank sources – these validation studies are currently underway.

Furthermore, despite their role as diagnostic biomarkers, one question which remains to be answered is whether or not the tumor is the main source of production of SAA1, HP and LCN15. Our *in silico* screen (**Fig. 7**) showed that all the three genes are upregulated in the kidney tumor compared to normal tissue. However, the kidney is not the main site of production of these proteins. According to the human protein atlas, SAA1 and HP are predominately produced in the liver and LCN15 is predominantly produced in the gastrointestinal tract [123]. It is possible that circulating proteins end up in the urine due to a disruption of the filtration barrier due to tissue damage from the growing tumor. One line of evidence speaking against this hypothesis is that we would then expect many more general serum proteins to score as high-confidence diagnostic biomarkers in our mass spectrometry data. Lastly, it also remains to be seen if SAA1, HP and LCN15 actually play a role in the tumorigenesis of ccRCC, or solely act as diagnostic biomarkers.

5. Conclusion and outlook

As the incidence of RCCs is increasing, so does the need for better diagnostic and prognostic biomarkers. The goal of my PhD was to tackle these questions and attempt to elucidate new biomarkers, specifically for ccRCC. In the present studies we have characterized CLEC18A as a correlate of tumor progression in ccRCC and shown that high *CLEC18A* expression is a favorable prognostic marker for patients. Furthermore, we have also characterized three urinary proteins (SAA1, HP and LCN15) as putative early-stage diagnostic biomarkers in a small-scale clinical discovery cohort. Despite these advances, many questions remain to be answered.

For CLEC18A, despite a role in hindering tumor growth in ccRCC, the mechanisms by which this is done remain to be elucidated. Further, the CLEC18 family of proteins is poorly characterized and the role of CLEC18 in normal physiology also remains to be discovered. Generation of *Clec18a* knock-out mice would provide a quintessential tool to answer these questions. Lastly, having identified a prognostic and beneficial role of CLEC18A in the tumors, it remains to be seen if CLEC18A levels has value as a prognostic or treatment marker in a clinical setting.

For urine-based diagnosis of ccRCC, the next immediate steps that must be performed are validations in larger, independent patient cohorts to independently assess the clinical utility of SAA1, HP and LCN15 levels in the urine of patients, in particular at early stages of disease. These studies are currently underway and will hopefully provide positive results in the near future. Further, to increase the efficiency and time investment of quantifying the three protein markers, developing monoclonal antibodies targeting the three diagnostic proteins of interest for standardized ELISA measurements would be of high interest.

In conclusion, this thesis provides new insights into how C-type lectins may influence tumor biology and highlights the promise of urinary biomarkers for early kidney cancer detection. Continued investigation into these areas could be clinically beneficial for patients.

6. References

1. Scholz, H., et al., *Kidney physiology and susceptibility to acute kidney injury: implications for renoprotection.* Nature Reviews Nephrology, 2021. **17**(5): p. 335-349.
2. Lynch, C.F. and M.B. Cohen, *Urinary system.* Cancer, 1995. **75**(S1): p. 316-329.
3. Finco, D.R., *Kidney function*, in *Clinical biochemistry of domestic animals*. 1997, Elsevier. p. 441-484.
4. Bakhshulloevna, S.D., *Functional Morphology of the Kidney.* American Journal of Pediatric Medicine and Health Sciences (2993-2149), 2023. **1**(4): p. 162-168.
5. Falkson, S.R. and B. Bordoni, *Anatomy, Abdomen and Pelvis, Bowman Capsule.* 2020.
6. Rehberg, P.B., *Studies on kidney function: the rate of filtration and reabsorption in the human kidney.* Biochemical Journal, 1926. **20**(3): p. 447.
7. Milne, M., *Tubular reabsorption and secretion.* J. Clin. Pathol, 1965. **18**: p. 515-519.
8. Pollak, M.R., et al., *The glomerulus: the sphere of influence.* Clinical Journal of the American Society of Nephrology, 2014. **9**(8): p. 1461-1469.
9. Yamada, E., *The fine structure of the renal glomerulus of the mouse.* The Journal of Cell Biology, 1955. **1**(6): p. 551-566.
10. Keener, J. and J. Sneyd, *Mathematical physiology 1: Cellular physiology*. 2009, Springer New York, NY, USA.
11. AC, G., *Urine formation by the kidneys: I. Glomerular filtration, renal blood flow, and their control.* Textbook of Medical Physiology, 1996.
12. Pei, L., et al., *Paracellular epithelial sodium transport maximizes energy efficiency in the kidney.* The Journal of clinical investigation, 2016. **126**(7): p. 2509-2518.
13. Aronson, P.S., *Ion exchangers mediating NaCl transport in the renal proximal tubule.* Cell biochemistry and biophysics, 2002. **36**: p. 147-153.
14. Curthoys, N.P. and O.W. Moe, *Proximal tubule function and response to acidosis.* Clinical Journal of the American Society of Nephrology, 2014. **9**(9): p. 1627-1638.
15. Wang, T., *Flow-activated transport events along the nephron.* Current opinion in nephrology and hypertension, 2006. **15**(5): p. 530-536.
16. Dantzler, W.H., et al., *Urine-concentrating mechanism in the inner medulla: function of the thin limbs of the loops of Henle.* Clinical Journal of the American Society of Nephrology, 2014. **9**(10): p. 1781-1789.
17. Imai, M., J. Taniguchi, and K. Tabei, *Function of thin loops of Henle.* Kidney international, 1987. **31**(2): p. 565-579.

18. Mount, D.B., *Thick ascending limb of the loop of Henle.* Clinical Journal of the American Society of Nephrology, 2014. **9**(11): p. 1974-1986.
19. Pannabecker, T.L., *Structure and function of the thin limbs of the loop of Henle.* Comprehensive physiology, 2012. **2**(3): p. 2063-2086.
20. Ko, B., et al., *A new model of the distal convoluted tubule.* American Journal of Physiology-Renal Physiology, 2012. **303**(5): p. F700-F710.
21. Subramanya, A.R. and D.H. Ellison, *Distal convoluted tubule.* Clinical Journal of the American Society of Nephrology, 2014. **9**(12): p. 2147-2163.
22. Purbhoo, K., *Physiology of the distal convoluted tubule and collecting duct.* Southern African Journal of Anaesthesia and Analgesia, 2020. **26**(6): p. S137-141.
23. Sasaki, S., et al., *Water channels in the kidney collecting duct.* Kidney international, 1995. **48**(4): p. 1082-1087.
24. Hsieh, J.J., et al., *Renal cell carcinoma.* Nature reviews Disease primers, 2017. **3**(1): p. 1-19.
25. Arjumand, W. and S. Sultana, *Role of VHL gene mutation in human renal cell carcinoma.* Tumor Biology, 2012. **33**: p. 9-16.
26. Rathmell, W.K. and S. Chen, *VHL inactivation in renal cell carcinoma: implications for diagnosis, prognosis and treatment.* Expert review of anticancer therapy, 2008. **8**(1): p. 63-73.
27. Algaba, F., et al., *Current pathology keys of renal cell carcinoma.* European urology, 2011. **60**(4): p. 634-643.
28. Muglia, V.F. and A. Prando, *Renal cell carcinoma: histological classification and correlation with imaging findings.* Radiologia brasileira, 2015. **48**: p. 166-174.
29. DeCastro, G.J. and J.M. McKiernan, *Epidemiology, clinical staging, and presentation of renal cell carcinoma.* Urologic Clinics of North America, 2008. **35**(4): p. 581-592.
30. Akhtar, M., I.A. Al-Bozom, and T. Al Hussain, *Papillary renal cell carcinoma (PRCC): an update.* Advances in anatomic pathology, 2019. **26**(2): p. 124-132.
31. Network, C.G.A.R., *Comprehensive molecular characterization of papillary renal-cell carcinoma.* New England Journal of Medicine, 2016. **374**(2): p. 135-145.
32. Amin, M.B., et al., *Chromophobe renal cell carcinoma: histomorphologic characteristics and evaluation of conventional pathologic prognostic parameters in 145 cases.* The American journal of surgical pathology, 2008. **32**(12): p. 1822-1834.
33. Garje, R., et al., *Comprehensive review of chromophobe renal cell carcinoma.* Critical reviews in oncology/hematology, 2021. **160**: p. 103287.

34. Gleeson, J.P., et al., *Comprehensive molecular characterization and response to therapy in fumarate hydratase–deficient renal cell carcinoma.* Clinical Cancer Research, 2021. **27**(10): p. 2910-2919.
35. Argani, P. *MiT family translocation renal cell carcinoma*. in *Seminars in diagnostic pathology*. 2015. Elsevier.
36. Tokuda, N., et al., *Collecting duct (Bellini duct) renal cell carcinoma: a nationwide survey in Japan.* The Journal of urology, 2006. **176**(1): p. 40-43.
37. Elliott, A. and E. Bruner, *Renal medullary carcinoma.* Archives of Pathology & Laboratory Medicine, 2019. **143**(12): p. 1556-1561.
38. Zhao, M., X.-l. He, and X.-d. Teng, *Mucinous tubular and spindle cell renal cell carcinoma: a review of clinicopathologic aspects.* Diagnostic pathology, 2015. **10**: p. 1-7.
39. Davis, C.F., et al., *The somatic genomic landscape of chromophobe renal cell carcinoma.* Cancer cell, 2014. **26**(3): p. 319-330.
40. Ricketts, C.J., et al., *The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma.* Cell reports, 2018. **23**(1): p. 313-326. e5.
41. Sato, Y., et al., *Integrated molecular analysis of clear-cell renal cell carcinoma.* Nature genetics, 2013. **45**(8): p. 860-867.
42. Borcherding, N., et al., *Mapping the immune environment in clear cell renal carcinoma by single-cell genomics.* Communications biology, 2021. **4**(1): p. 122.
43. Jonsson, G., et al., *Urinary multi-omics reveal non-invasive diagnostic biomarkers in clear cell renal cell carcinoma.* bioRxiv, 2024: p. 2024.08. 12.607453.
44. Giaccia, A., B.G. Siim, and R.S. Johnson, *HIF-1 as a target for drug development.* Nature reviews Drug discovery, 2003. **2**(10): p. 803-811.
45. Lonser, R.R., et al., *von Hippel-Lindau disease.* The Lancet, 2003. **361**(9374): p. 2059-2067.
46. Maxwell, P.H., et al., *The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis.* Nature, 1999. **399**(6733): p. 271-275.
47. Hoefflin, R., et al., *HIF-1α and HIF-2α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice.* Nature communications, 2020. **11**(1): p. 4111.
48. Kaluz, S., M. Kaluzová, and E.J. Stanbridge, *Regulation of gene expression by hypoxia: integration of the HIF-transduced hypoxic signal at the hypoxia-responsive element.* Clinica Chimica Acta, 2008. **395**(1-2): p. 6-13.

49. Li, L., S.M. Hossain, and M.R. Eccles, *The Role of the PAX Genes in Renal Cell Carcinoma.* International Journal of Molecular Sciences, 2024. **25**(12): p. 6730.
50. Luu, V.-D., et al., *Loss of VHL and hypoxia provokes PAX2 up-regulation in clear cell renal cell carcinoma.* Clinical Cancer Research, 2009. **15**(10): p. 3297-3304.
51. Keith, B., R.S. Johnson, and M.C. Simon, *HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression.* Nature Reviews Cancer, 2012. **12**(1): p. 9-22.
52. Downes, N.L., et al., *Differential but complementary HIF1α and HIF2α transcriptional regulation.* Molecular Therapy, 2018. **26**(7): p. 1735-1745.
53. Patel, S.A., et al., *The renal lineage factor PAX8 controls oncogenic signalling in kidney cancer.* Nature, 2022. **606**(7916): p. 999-1006.
54. Hakimi, A.A., et al., *Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma.* European urology, 2013. **63**(5): p. 848-854.
55. Ingels, A., et al., *Complementary roles of surgery and systemic treatment in clear cell renal cell carcinoma.* Nature Reviews Urology, 2022. **19**(7): p. 391-418.
56. Hu, X., et al., *Radical versus partial nephrectomy for T1 non-clear cell renal cell carcinoma.* European Journal of Surgical Oncology, 2023. **49**(8): p. 1519-1523.
57. Powles, T., et al., *Pembrolizumab versus placebo as post-nephrectomy adjuvant therapy for clear cell renal cell carcinoma (KEYNOTE-564): 30-month follow-up analysis of a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial.* The Lancet Oncology, 2022. **23**(9): p. 1133-1144.
58. Choueiri, T.K., et al., *Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma.* New England Journal of Medicine, 2021. **385**(8): p. 683-694.
59. Choueiri, T.K., et al., *Overall survival with adjuvant pembrolizumab in renal-cell carcinoma.* New England Journal of Medicine, 2024. **390**(15): p. 1359-1371.
60. Barata, P.C. and B.I. Rini, *Treatment of renal cell carcinoma: current status and future directions.* CA: a cancer journal for clinicians, 2017. **67**(6): p. 507-524.
61. Janisch, F., et al., *MP25-04 OUTCOMES OF METASTATIC RENAL CELL CARCINOMA PATIENTS WITH SARCOMATOID HISTOLOGY-A SINGLE TERTIARY CARE CENTER STUDY IN THE TKI ERA.* Journal of Urology, 2019. **201**: p. e343-e344.
62. Choueiri, T.K., et al., *Cabozantinib versus everolimus in advanced renal-cell carcinoma.* New England Journal of Medicine, 2015. **373**(19): p. 1814-1823.

63. Chow, L.Q. and S.G. Eckhardt, *Sunitinib: from rational design to clinical efficacy.* Journal of clinical oncology, 2007. **25**(7): p. 884-896.
64. Escudier, B., et al., *Sorafenib in advanced clear-cell renal-cell carcinoma.* New England Journal of Medicine, 2007. **356**(2): p. 125-134.
65. Motzer, R.J., et al., *Pazopanib versus sunitinib in metastatic renal-cell carcinoma.* New England Journal of Medicine, 2013. **369**(8): p. 722-731.
66. Buczek, M., et al., *Resistance to tyrosine kinase inhibitors in clear cell renal cell carcinoma: from the patient's bed to molecular mechanisms.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2014. **1845**(1): p. 31-41.
67. F Bielecka, Z., et al., *Mechanisms of acquired resistance to tyrosine kinase inhibitors in clear-cell renal cell carcinoma (ccRCC).* Current signal transduction therapy, 2013. **8**(3): p. 219-228.
68. Gebrael, G., et al., *Update on combined immunotherapy for the treatment of advanced renal cell carcinoma.* Human Vaccines & Immunotherapeutics, 2023. **19**(1): p. 2193528.
69. Albiges, L., et al., *Pembrolizumab plus lenvatinib as first-line therapy for advanced non-clear-cell renal cell carcinoma (KEYNOTE-B61): A single-arm, multicentre, phase 2 trial.* The Lancet Oncology, 2023. **24**(8): p. 881-891.
70. Bolan, P., et al., *Real world evidence comparison of first-line (1L) immune-oncology (IO)/tyrosine kinase inhibitor (TKI) vs. IO/IO combination therapy in renal cell carcinoma (RCC)*. 2024, American Society of Clinical Oncology.
71. Vano, Y.-A., et al., *Nivolumab, nivolumab–ipilimumab, and VEGFR-tyrosine kinase inhibitors as first-line treatment for metastatic clear-cell renal cell carcinoma (BIONIKK): A biomarker-driven, open-label, non-comparative, randomised, phase 2 trial.* The Lancet Oncology, 2022. **23**(5): p. 612-624.
72. Tykodi, S.S., et al., *First-line pembrolizumab (pembro) monotherapy in advanced clear cell renal cell carcinoma (ccRCC): Updated results for KEYNOTE-427 cohort A*. 2019, American Society of Clinical Oncology.
73. Vaishampayan, U., et al., *Avelumab monotherapy as first-line or second-line treatment in patients with metastatic renal cell carcinoma: phase Ib results from the JAVELIN Solid Tumor trial.* Journal for ImmunoTherapy of Cancer, 2019. **7**: p. 1-9.
74. Atkins, M.B., et al., *Phase II study of nivolumab and salvage nivolumab/ipilimumab in treatment-naive patients with advanced clear cell renal cell carcinoma (HCRN GU16-260-Cohort A).* Journal of Clinical Oncology, 2022. **40**(25): p. 2913-2923.

75. Deleuze, A., et al., *Immunotherapy in renal cell carcinoma: the future is now.* International Journal of Molecular Sciences, 2020. **21**(7): p. 2532.
76. Dutcher, J.P., et al., *On the shoulders of giants: the evolution of renal cell carcinoma treatment—cytokines, targeted therapy, and immunotherapy.* American Society of Clinical Oncology Educational Book, 2020. **40**: p. 418-435.
77. Ballesteros, P.Á., et al., *Molecular mechanisms of resistance to immunotherapy and antiangiogenic treatments in clear cell renal cell carcinoma.* Cancers, 2021. **13**(23): p. 5981.
78. June, C.H., et al., *CAR T cell immunotherapy for human cancer.* Science, 2018. **359**(6382): p. 1361-1365.
79. Pal, S.K., et al., *CD70-Targeted Allogeneic CAR T-Cell Therapy for Advanced Clear Cell Renal Cell Carcinoma.* Cancer Discovery, 2024: p. OF1-OF14.
80. Wang, Y., et al., *Evolution of cell therapy for renal cell carcinoma.* Molecular Cancer, 2024. **23**(1): p. 8.
81. Bukavina, L., et al., *Epidemiology of Renal Cell Carcinoma: 2022 Update.* Eur Urol, 2022. **82**(5): p. 529-542.
82. Gray, R.E. and G.T. Harris, *Renal cell carcinoma: diagnosis and management.* American family physician, 2019. **99**(3): p. 179-184.
83. Sankineni, S., et al. *Imaging of renal cell carcinoma*. in *Urologic Oncology: Seminars and Original Investigations*. 2016. Elsevier.
84. Vogel, C., et al., *Imaging in suspected renal-cell carcinoma: systematic review.* Clinical genitourinary cancer, 2019. **17**(2): p. e345-e355.
85. Delahunt, B., et al., *Grading of renal cell carcinoma.* Histopathology, 2019. **74**(1): p. 4-17.
86. Marostica, E., et al., *Development of a histopathology informatics pipeline for classification and prediction of clinical outcomes in subtypes of renal cell carcinoma.* Clinical Cancer Research, 2021. **27**(10): p. 2868-2878.
87. Alcala, K., et al., *Kidney function and risk of Renal Cell Carcinoma.* Cancer Epidemiology, Biomarkers & Prevention, 2023. **32**(11): p. 1644-1650.
88. Shingarev, R. and E.A. Jaimes, *Renal cell carcinoma: new insights and challenges for a clinician scientist.* American Journal of Physiology-Renal Physiology, 2017. **313**(2): p. F145-F154.

89. Choi, J.W., et al., *Preoperative prediction of the stage, size, grade, and necrosis score in clear cell renal cell carcinoma using MRI-based radiomics.* Abdominal Radiology, 2021. **46**: p. 2656-2664.
90. Bodalal, Z., et al., *Radiogenomics: bridging imaging and genomics.* Abdominal radiology, 2019. **44**(6): p. 1960-1984.
91. Khaleel, S., et al., *Radiogenomics in clear cell renal cell carcinoma: A review of the current status and future directions.* Cancers, 2022. **14**(9): p. 2085.
92. Brooks, S.A., et al., *ClearCode34: a prognostic risk predictor for localized clear cell renal cell carcinoma.* European urology, 2014. **66**(1): p. 77-84.
93. Ghatalia, P. and W.K. Rathmell, *Systematic review: clearcode 34–a validated prognostic signature in clear cell renal cell carcinoma (ccRCC).* Kidney cancer, 2018. **2**(1): p. 23-29.
94. Kim, H.L., et al., *Using protein expressions to predict survival in clear cell renal carcinoma.* Clinical cancer research, 2004. **10**(16): p. 5464-5471.
95. Harpole, M., J. Davis, and V. Espina, *Current state of the art for enhancing urine biomarker discovery.* Expert review of Proteomics, 2016. **13**(6): p. 609-626.
96. Ploussard, G. and A. De La Taille, *Urine biomarkers in prostate cancer.* Nature Reviews Urology, 2010. **7**(2): p. 101-109.
97. Yang, Y., et al., *Excavation of diagnostic biomarkers and construction of prognostic model for clear cell renal cell carcinoma based on urine proteomics.* Frontiers in Oncology, 2023. **13**: p. 1170567.
98. Jones, H.B., *III. On a new substance occurring in the urine of a patient with mollities ossium.* Philosophical Transactions of the Royal Society of London, 1848(138): p. 55-62.
99. Witjes, J.A., et al., *EAU-ESMO Consensus statements on the management of advanced and variant bladder cancer—an international collaborative multistakeholder effort†: under the auspices of the EAU-ESMO Guidelines Committees.* European urology, 2020. **77**(2): p. 223-250.
100. Hirasawa, Y., et al., *Diagnostic performance of Oncuria™, a urinalysis test for bladder cancer.* Journal of Translational Medicine, 2021. **19**: p. 1-10.
101. Mandrekar, J.N., *Receiver operating characteristic curve in diagnostic test assessment.* Journal of Thoracic Oncology, 2010. **5**(9): p. 1315-1316.

102. Maas, M., T. Todenhöfer, and P.C. Black, *Urine biomarkers in bladder cancer—Current status and future perspectives.* Nature Reviews Urology, 2023. **20**(10): p. 597-614.
103. Eich, M.-L., et al., *Incidence and distribution of UroSEEK gene panel in a multi-institutional cohort of bladder urothelial carcinoma.* Modern Pathology, 2019. **32**(10): p. 1544-1550.
104. Springer, S.U., et al., *Non-invasive detection of urothelial cancer through the analysis of driver gene mutations and aneuploidy.* Elife, 2018. **7**: p. e32143.
105. Dudley, J.C., et al., *Detection and surveillance of bladder cancer using urine tumor DNA.* Cancer discovery, 2019. **9**(4): p. 500-509.
106. Chaudhuri, A.A., et al., *Emerging roles of urine-based tumor DNA analysis in bladder cancer management.* JCO Precision Oncology, 2020. **4**: p. 806-817.
107. Ferro, M., et al., *Liquid biopsy biomarkers in urine: a route towards molecular diagnosis and personalized medicine of bladder cancer.* Journal of personalized medicine, 2021. **11**(3): p. 237.
108. Sandim, V., et al. *Proteomic analysis reveals differentially secreted proteins in the urine from patients with clear cell renal cell carcinoma*. in *Urologic Oncology: Seminars and Original Investigations*. 2016. Elsevier.
109. Santorelli, L., et al., *In-depth mapping of the urinary N-glycoproteome: distinct signatures of ccRCC-related progression.* Cancers, 2020. **12**(1): p. 239.
110. Chinello, C., et al., *Proteomics of liquid biopsies: Depicting RCC infiltration into the renal vein by MS analysis of urine and plasma.* Journal of proteomics, 2019. **191**: p. 29-37.
111. Di Meo, A., et al., *Searching for prognostic biomarkers for small renal masses in the urinary proteome.* International journal of cancer, 2020. **146**(8): p. 2315-2325.
112. Hu, R., et al., *Bioplasmonic paper–based assay for perilipin-2 non-invasively detects renal cancer.* Kidney international, 2019. **96**(6): p. 1417-1421.
113. Morrissey, J.J., et al., *Evaluation of urine aquaporin-1 and perilipin-2 concentrations as biomarkers to screen for renal cell carcinoma: a prospective cohort study.* JAMA oncology, 2015. **1**(2): p. 204-212.
114. Outeiro-Pinho, G., et al., *MicroRNA-30a-5p me: a novel diagnostic and prognostic biomarker for clear cell renal cell carcinoma in tissue and urine samples.* Journal of Experimental & Clinical Cancer Research, 2020. **39**: p. 1-11.

115. Oto, J., et al., *Urinary microRNAs: looking for a new tool in diagnosis, prognosis, and monitoring of renal cancer.* Current urology reports, 2020. **21**: p. 1-8.
116. Nuzzo, P.V., et al., *Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes.* Nature medicine, 2020. **26**(7): p. 1041-1043.
117. Zhang, Y., et al., *Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response.* Proceedings of the National Academy of Sciences, 2021. **118**(24): p. e2103240118.
118. Burg, M., et al., *Organic solutes in fluid absorption by renal proximal convoluted tubules.* American Journal of Physiology-Legacy Content, 1976. **231**(2): p. 627-637.
119. Weinstein, J.N., et al., *The cancer genome atlas pan-cancer analysis project.* Nature genetics, 2013. **45**(10): p. 1113-1120.
120. Grossman, R.L., et al., *Toward a shared vision for cancer genomic data.* New England Journal of Medicine, 2016. **375**(12): p. 1109-1112.
121. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**: p. 1-21.
122. Thul, P.J., et al., *A subcellular map of the human proteome.* Science, 2017. **356**(6340): p. eaal3321.
123. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
124. Uhlén, M., et al., *The human secretome.* Science signaling, 2019. **12**(609): p. eaaz0274.
125. Cummings, R.D., et al., *C-type lectins.* Essentials of Glycobiology [Internet]. 4th edition, 2022.
126. Plato, A., J.A. Willment, and G.D. Brown, *C-type lectin-like receptors of the dectin-1 cluster: ligands and signaling pathways.* International reviews of immunology, 2013. **32**(2): p. 134-156.
127. Mnich, M.E., R. Van Dalen, and N.M. Van Sorge, *C-type lectin receptors in host defense against bacterial pathogens.* Frontiers in cellular and infection microbiology, 2020. **10**: p. 309.
128. Hoving, J.C., G.J. Wilson, and G.D. Brown, *Signalling C-type lectin receptors, microbial recognition and immunity.* Cellular microbiology, 2014. **16**(2): p. 185-194.
129. Hoffmann, D., et al., *Identification of lectin receptors for conserved SARS-CoV-2 glycosylation sites.* The EMBO Journal, 2021. **40**(19): p. e108375.
130. Seifert, L., et al., *The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression.* Nature, 2016. **532**(7598): p. 245-249.

131. Daley, D., et al., *Dectin 1 activation on macrophages by galectin 9 promotes pancreatic carcinoma and peritumoral immune tolerance.* Nature medicine, 2017. **23**(5): p. 556-567.
132. Ding, D., et al., *C-type lectins facilitate tumor metastasis.* Oncology letters, 2017. **13**(1): p. 13-21.
133. Kimura, Y., et al., *The innate immune receptor Dectin-2 mediates the phagocytosis of cancer cells by Kupffer cells for the suppression of liver metastasis.* Proceedings of the National Academy of Sciences, 2016. **113**(49): p. 14097-14102.
134. Seifert, L., et al., *Dectin-1 regulates hepatic fibrosis and hepatocarcinogenesis by suppressing TLR4 signaling pathways.* Cell reports, 2015. **13**(9): p. 1909-1921.
135. Lanier, L.L., *NKG2D Receptor and Its Ligands in Host Defense.* Cancer Immunology Research, 2015. **3**(6): p. 575-582.
136. Jonsson, G., et al., *CLEC18A interacts with sulfated GAGs and controls clear cell renal cell carcinoma progression.* bioRxiv, 2024: p. 2024.07. 08.602586.
137. Hashmi, M. and F. Limaiem, *Renal clear cell cancer.* StatPearls, 2023.
138. Yang, J., K. Wang, and Z. Yang, *Treatment strategies for clear cell renal cell carcinoma: Past, present and future.* Frontiers in Oncology, 2023. **13**: p. 1133832.
139. Moreira, M., et al., *Resistance to cancer immunotherapy in metastatic renal cell carcinoma.* Cancer Drug Resistance, 2020. **3**(3): p. 454.
140. Alix-Panabières, C. and K. Pantel, *Liquid biopsy: from discovery to clinical application.* Cancer discovery, 2021. **11**(4): p. 858-873.
141. Crowley, E., et al., *Liquid biopsy: monitoring cancer-genetics in the blood.* Nature reviews Clinical oncology, 2013. **10**(8): p. 472-484.
142. Zieren, R.C., et al., *Diagnostic liquid biopsy biomarkers in renal cell cancer.* Nature Reviews Urology, 2024. **21**(3): p. 133-157.
143. Huang, Y.-L., et al., *Human CLEC18 gene cluster contains C-type lectins with differential glycan-binding specificity.* Journal of Biological Chemistry, 2015. **290**(35): p. 21252-21263.
144. Chang, C.-M., W.-C. Chang, and S.L. Hsieh, *Characterization of the genetic variation and evolutionary divergence of the CLEC18 family.* Journal of Biomedical Science, 2024. **31**(1): p. 53.
145. Liao, T.-L., et al., *Association of C-type lectin 18 levels with extrahepatic manifestations in chronic HCV infection.* Scientific Reports, 2018. **8**(1): p. 17287.

146. Tsai, T.-Y., et al., *The human C-type lectin 18 is a potential biomarker in patients with chronic hepatitis B virus infection.* Journal of biomedical science, 2018. **25**: p. 1-10.
147. Hsu, Y.-W., et al., *Human rs75776403 polymorphism links differential phenotypic and clinical outcomes to a CLEC18A p. T151M-driven multiomics.* Journal of Biomedical Science, 2022. **29**(1): p. 43.
148. Mathuru, A.S. and S. Jesuthasan, *The medial habenula as a regulator of anxiety in adult zebrafish.* Frontiers in neural circuits, 2013. **7**: p. 99.
149. McLaughlin, I., J.A. Dani, and M. De Biasi, *The medial habenula and interpeduncular nucleus circuitry is critical in addiction, anxiety, and mood regulation.* Journal of neurochemistry, 2017. **142**: p. 130-143.
150. Viswanath, H., et al., *The medial habenula: still neglected.* Frontiers in human neuroscience, 2014. **7**: p. 931.
151. Gasteiger, E., *Protein identification and analysis tools on the ExPASy server*. 2005, The Proteomics Protocols Handbook/Humana Press.
152. Qi, X., et al., *The uniqueness of clear cell renal cell carcinoma: summary of the process and abnormality of glucose metabolism and lipid metabolism in ccRCC.* Frontiers in Oncology, 2021. **11**: p. 727778.
153. Yong, C., G.D. Stewart, and C. Frezza, *Oncometabolites in renal cancer.* Nature Reviews Nephrology, 2020. **16**(3): p. 156-172.
154. Bader, J.M., V. Albrecht, and M. Mann, *MS-based proteomics of body fluids: the end of the beginning.* Molecular & Cellular Proteomics, 2023. **22**(7).
155. Gillette, M.A., C.R. Jimenez, and S.A. Carr, *Clinical Proteomics: A Promise Becoming Reality.* Molecular & Cellular Proteomics, 2024. **23**(2).
156. Tabatabaei, M.S. and M. Ahmed, *Enzyme-linked immunosorbent assay (ELISA)*, in *Cancer cell biology: Methods and protocols*. 2022, Springer. p. 115-134.
157. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics.* Molecular & cellular proteomics, 2012. **11**(11): p. 1475-1488.

7. Proof of submission

7.1 Publication 1 proof of submission

From: commsbio@nature.com
Subject: Receipt of open access submission COMMSBIO-24-5701: please review our policies
Date: 10. September 2024 at 17:17
To: gustav.jonsson@imba.oeaw.ac.at

Dear Mr Jonsson,

Thank you for submitting your Article entitled "CLEC18A interacts with sulfated GAGs and controls clear cell renal cell carcinoma progression" to Communications Biology.

Please be aware that Communications Biology is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit our open access page. Additional information about open access at the Nature Portfolio can be found here.

You may check the status of your submission by selecting the "Check manuscript status" link at the following URL:

https://mts-commsbio.nature.com/cgi-bin/main.plex?el=A3Cx4KRc3A7BtxT4F6A9ftdgvDPCcWUH96Wu5Y9NEpNAAZ

Please note that submission to Communications Biology is taken to imply that there is no significant overlap between the submitted manuscript and any other papers from the same authors under consideration or in press elsewhere. (Abstracts or unrefereed web preprints do not compromise novelty.) Please email our office copies of any related manuscripts with overlap in authorship that are under consideration or in press elsewhere.

Our data availability policy expresses our preference for research data to be archived in public, specialised data repositories where applicable, and we also encourage our authors to consider submitting an additional publication to Scientific Data alongside their primary research article. Scientific Data publishes Data Descriptors, which are peer-reviewed articles enabling authors to provide comprehensive methodological detail about their data. Publishing a Data Descriptor can facilitate discovery, reuse and citation of your research data. Please contact the editorial team of Scientific Data at scientificdata@nature.com if you are interested in publishing a Data Descriptor supporting your current manuscript.

Furthermore, Articles and Reviews published in Communications Biology must contain a declaration of competing financial and non-financial interests. Please see our policy for more information.

Communications Biology uses a transparent peer review system, in which we publish the reviewer comments to the authors and author rebuttal letters of our research articles online as a supplementary peer review file. Our authors are given the opportunity to opt out of transparent peer review at the point of acceptance. Furthermore, on author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication.

To increase transparency in the peer review process, Springer Nature has partnered with Research Square to provide you with a private author dashboard where you can access more detailed information on the status of your manuscript throughout peer review via a peer review timeline. Research Square will be in touch separately with more details on this service.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature Portfolio journals. Please check your account regularly and ensure that we have your current contact information.

In addition, Nature Portfolio journals encourage all authors and reviewers to associate an Open Researcher and Contributor Identifier (ORCID) to their account. ORCID is a community-based initiative that provides an open, non-proprietary and transparent registry of unique identifiers to help disambiguate research contributions.

Thank you for submitting your work to Communications Biology. We will contact you shortly with a decision.

Open access

Communications Biology is an open access journal. This means your article will be free to read online if it is accepted for publication. There is a cost associated with publishing open access called an article processing charge (APC). Check the APC for Communications Biology here:
https://www.springernature.com/gp/open-research/journals-books/journals

The APC for your article may be covered by one of Springer Nature's fully open access agreements. We will check this if your article is accepted.

View our institutional agreements:
https://www.springernature.com/gp/open-research/institutional-agreements

There may also be other sources of funding available to you. You can get help to find and apply for funding through our free Funding and Support Service:
https://www.springernature.com/gp/open-research/funding

Best regards,

Manuscript Administration
Communications Biology

This email has been sent through the Springer Nature Tracking System NY-610A-NPG&MTS

Confidentiality Statement:

This e-mail is confidential and subject to copyright. Any unauthorised use or disclosure of its contents is prohibited. If you have received this email in error please notify our Manuscript Tracking System Helpdesk team at http://platformsupport.nature.com .

Details of the confidentiality and pre-publicity policy may be found here http://www.nature.com/authors/policies/confidentiality.html

Privacy Policy I Update Profile

7.2 Publication 2 proof of submission

From: scitranslmededitors@aaas.org
Subject: Successful Submission of a Manuscript to Science Translational Medicine (adt0728)
Date: 11. September 2024 at 19:20
To: gustav.jonsson@imba.oeaw.ac.at

Science Translational Medicine

Manuscript Title: Urinary multi-omics reveal non-invasive diagnostic biomarkers in clear cell renal cell carcinoma
Author: Jonsson
Manuscript Number: adt0728

Dear Dr. Jonsson,

You are listed as a coauthor on the above manuscript, which has recently been submitted to Science Translational Medicine. According to Science Translational Medicine policy, all autho must have seen and approved the submission of their manuscript. Please click on the followi link to confirm your authorship on this manuscript. The link also allows you to complete your demographic information, and to update your subject-area expertise and institutions.

https://cts.sciencemag.org/scc/#/action/article/67a38f90-3997-45c1-b46d-621e72f65ba9/author/4574457/verifyAuthorship?email=gustav.jonsson@imba.oeaw.ac.at&token=26ad1d4a-7062-11ef-869b-0ad102f82f1d

If you have not read this paper or do not approve its submission to Science Translational Medicine, please let us know as soon as possible. Please refer to the manuscript number list above in any correspondence (you can just reply to this message).

You can see the status of your manuscript at any time by logging into your account at the Science Journals Content Tracking System at https://cts.sciencemag.org.

Once you login, we encourage you to link your account to your ORCID ID, an identifier that facilitates the correct attribution of your publications to you. To learn more about ORCID or to obtain an ORCID ID, visit their site at: http://orcid.org.

Your manuscript is now undergoing an initial screening to determine whether it will be sent fo in-depth review.

Sincerely,
The Editors